Sweden's scientists spar over its relaxed pandemic policies p. 159

Black phosphorus anodes for Li ion batteries p. 192

Intersexuality in female moles p. 208

Science

$15
9 OCTOBER 2020
sciencemag.org

AAAS

SPIKE IN MOTION

SARS-CoV-2 surface protein has flexible hinges p. 203

CONTENTS

9 OCTOBER 2020 • VOLUME 370 • ISSUE 6513

As COVID-19 cases climbed in May, Stockholm remained busy—and largely maskless.

PHOTO: BLOOMBERG/GETTY IMAGES

NEWS

IN BRIEF

150 News at a glance

IN DEPTH

153 Astronomy is—and has—a climate problem
Big carbon footprints and rising vulnerabilities push astronomers to take action *By D. Clery*

154 A call to test new vaccines head to head, in monkeys
Proposed comparison study could clarify safety and which vaccines work best
By J. Cohen

156 With to-do list checked off, U.S. physicists ask, 'What's next?'
Giant neutrino experiment is only sure thing for the field *By A. Cho*

157 Newly found viruses suggest rubella originated in animals
Two relatives discovered in bats, mice, and zoo animals suggest "German measles" jumped the species barrier *By A. Gibbons*

158 NASA mission set to sample carbon-rich asteroid
Bennu's thick carbonate veins suggest past life as a water world *By P. Voosen*
10.1126/SCIENCE.ABC3660; 10.1126/SCIENCE.ABC3522; 10.1126/SCIENCE.ABC3557; 10.1126/SCIADV.ABC3550; 10.1126/SCIADV.ABD3649; 10.1126/SCIADV.ABC3699

FEATURES

159 Sweden's gamble
The country's pandemic policies came at a high price—and created painful rifts in its scientific community
By G. Vogel

INSIGHTS

PERSPECTIVES

164 Novel communities are a risky business
High turnover leads to novel combinations of species and involves high extinction
By M. Dornelas and J. S. Madin
REPORT p. 220

166 Making a bed for viral infections
An inflamed mucosal niche is permissive for symptomatic respiratory syncytial virus infection
By A. S. Mirchandani and S. R. Walmsley
RESEARCH ARTICLE p. 188

167 Inhibiting Ebola virus and SARS-CoV-2 entry
Screening identifies CD74 as a cellular antiviral protein against Ebola virus and SARS-CoV-2
By A. I. Wells and C. B. Coyne
REPORT p. 241

168 A new villain in neuronal death
A newly discovered partnership for glutamate triggers brain toxicity *By S. Jones*
RESEARCH ARTICLE p. 191

170 Layer-cake 2D superconductivity
An easy-to-synthesize quantum material may help unveil quantum effects *By L. M. Schoop*
REPORT p. 231

171 Better, faster, and even cheap
An inexpensive gold grid with submicrometer holes minimizes protein sample movement *By M. Rapp and B. Carragher*
REPORT p. 223

172 Contextualizing bats as viral reservoirs
Preventing zoonotic emergence from bats requires integrative research
By D. G. Streicker and A. T. Gilbert

POLICY FORUM

174 U.S. policy puts the safe development of space at risk
Promoting national regulation of space mining rather than multilateral governance risks a "race to the bottom" *By A. Boley and M. Byers*

BOOKS *ET AL.*

177 Plotting a path to a more sustainable future
Persistent political pressure staves off global warming in a new climate allegory
By C. Abbott

178 Moving beyond buzzwords
Rejecting conventional connotations, two books urge readers to rethink innovation
By D. Greenbaum and M. Gerstein

LETTERS

179 Retraction
By Y. Wu et al.

Published by AAAS

179 Hidden ethical costs of conservation
By K. E. Lynch and D. T. Blumstein

180 Pragmatic animal welfare is independent of feelings
By R. Arlinghaus et al.

180 Animal welfare science aids conservation
By J. O. Hampton et al.

181 Response
By N. Sekar and D. Shiller

181 Techincal comment abstracts

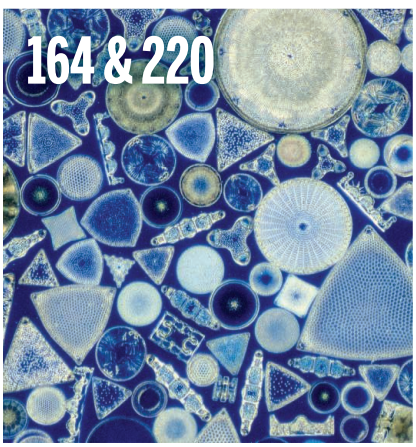


PHOTO: M. I. WALKER/SCIENCE SOURCE

RESEARCH

IN BRIEF

182 From *Science* and other journals

REVIEW

185 Biotechnology
Bioengineered human blood vessels
L. E. Niklason and J. H. Lawson
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAW8682

186 Electrochemistry
Recent advances in solid oxide cell technology for electrolysis *A. Hauch* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA6118

RESEARCH ARTICLES

187 HIV
Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro *D. E. Christensen* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABC8420

188 Viral infections
Neutrophilic inflammation in the respiratory mucosa predisposes to RSV infection *M. S. Habibi* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA9301
PERSPECTIVE p. 166

189 Structural biology
Conformational states dynamically populated by a kinase determine its function *T. Xie* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABC2754

190 Metallurgy
The heterogeneity of persistent slip band nucleation and evolution in metals at the micrometer scale *S. Lavenstein* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABB2690

191 Neuroscience
Coupling of NMDA receptors and TRPM4 guides discovery of unconventional neuroprotectants *J. Yan* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAY3302
PERSPECTIVE p. 168

192 Batteries
Black phosphorus composites with engineered interfaces for high-rate high-capacity lithium storage *H. Jin* et al.

198 Quantum gravity
Spacetime from bits *M. Van Raamsdonk*

203 Coronavirus
In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges
B. Turoňová et al.

208 Developmental biology
The mole genome reveals regulatory rearrangements associated with adaptive intersexuality
F. M. Real et al.

REPORTS

214 Organic chemistry
Electro-inductive effect: Electrodes as functional groups with tunable electronic properties *J. Heo* et al.

220 Novel communities
Increased extinction in the emergence of novel ecological communities
J. M. Pandolfi et al.
PERSPECTIVE p. 164

223 Electron microscopy
Cryo-EM with sub–1 Å specimen movement *K. Naydenova* et al.
PERSPECTIVE p. 171

227 Plant science
WUSCHEL triggers innate antiviral immunity in plant stem cells *H. Wu* et al.

231 Superconductivity
Clean 2D superconductivity in a bulk van der Waals superlattice *A. Devarakonda* et al.
PERSPECTIVE p. 170

237 Vaccines
Influenza vaccine–induced human bone marrow plasma cells decline within a year after vaccination *C. W. Davis* et al.

241 Coronavirus
MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses *A. Bruchez* et al.
PERSPECTIVE p. 167

247 Neuroscience
Alternating sequences of future and past behavior encoded within hippocampal theta oscillations *M. Wang* et al.

DEPARTMENTS

149 Editorial
Here today, gone tomorrow
By Peter H. Raven and Scott E. Miller

258 Working Life
A pandemic journey *By Sergio E. Ramos*

ON THE COVER

The spike protein (teal), covered in protective glycan molecules, on the surface of SARS-CoV-2 is the primary target for COVID-19 vaccine development. The spike contains three hinges that enable it to be highly flexible on the viral surface and may facilitate host cell entry. The spike positions shown here are from structural data of intact SARS-CoV-2 virions. See page 203. *Illustration: V. Altounian/*Science*; Data: Paul Ehrlich Institute, European Molecular Biology Laboratory, and Max Planck Institute of Biophysics*

Science Staff 148
New Products 254
Science Careers 255

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

EDITORIAL

Words matter

H. Holden Thorp
Editor-in-Chief, *Science* Journals.
hthorp@aaas.org;
@hholdenthorp

For more than 200 years, U.S. presidents have strived to deliver words of inspiration and humility that will stand the test of time. Even in a contentious debate in which the Oval Office is at stake, we expect a standing president to inspire and unite the country. Not surprisingly, President Donald Trump used the occasion of the first presidential debate of 2020 to deny reality, insult his opponent, and praise himself—all with his customary lean vocabulary.

Over the ages, U.S. presidents have pored over parchment, legal pads, and laptops to create the scripture of the nation: "…the only thing we have to fear is…fear itself," "a date which will live in infamy," "ask not what your country can do for you…," and "Mr. Gorbachev, tear down this wall" resonate ever stronger over time.

There are no such inspiring words today. President Trump has no interest in lifting up Americans or the rest of the world. He only wants to bring society down with his rhetorical carnage. During the past 4 years we've heard "very fine people, on both sides," "I need loyalty," and "shithole countries." And from the debate stage, he told one group of militant right-wing thugs to "stand back and stand by," as if he was asking them to cool it for now while awaiting further orders.

When it comes to the crisis of coronavirus disease 2019 (COVID-19), Trump's words could not be more destructive. When scientists tried to tell him a crisis was coming, he called it "their new hoax." About the extraordinary number of lives lost he says, "it is what it is." His plan for conquering the virus is simply that "like a miracle, it will disappear." His communications strategy is "I always wanted to play it down." And as for his role as the leader of a country in crisis, he says, "No, I don't take responsibility at all." It's safe to say that Presidents Roosevelt, Kennedy, and Reagan would have said something more comforting and profound.

His words are particularly painful for science. Long days and nights have been spent by scientists working at the bench, fighting to understand the causative virus and methods to defeat it. Epidemiologists have been analyzing their models trying to devise mitigations. Physician-scientists and their colleagues in academic hospitals have developed new approaches to bring down the death rate substantially. And through all this, these researchers rarely have heard one word of acknowledgment from their president. It's no wonder that the braggart who said, "I alone can fix it" can't bring himself to admit that he is not the person who will get us through the pandemic. Even with the vaccine that so many have worked toward, Trump has manufactured a rationale by which he is the one who deserves the credit.

Now that Trump himself has been diagnosed with COVID-19, he is seeing first-hand the benefits of the science he has long undermined. The experimental antibody cocktail that he is taking is the fruit of a gargantuan effort by scientists fighting hard to understand the viral spike protein and characteristics of the most potent neutralizing antibodies. That effort stands on the shoulders of decades of fundamental research in immunology and structural biology, science that Trump has insulted and devalued by cutting funding for the National Institutes of Health in every budget he has submitted, but even more so with his words. Words that criticized science for discouraging use of hydroxychloroquine, a drug that he is not taking now that he has COVID-19. Words that discouraged public health interventions that cost lives and led directly to his own superspreader event in the White House's Rose Garden. Words that painted career scientists at the Food and Drug Administration, who have devoted their lives to protecting the public's health, as the "deep state." And words that implied that the only reason scientists were working so hard was because of his compelling exhortations to do him a favor and speed it up. A speedy recovery is hoped for President Trump and everyone who is battling this disease, as science strives to provide an effective vaccine.

Those of us who live in Washington, DC, are surrounded by the words that define America. We can go on long walks to the World War II Memorial, along the Reflecting Pool, and up the steps of the Lincoln Memorial. We can stand where Dr. Martin Luther King Jr. said, "I have a dream." And we can see chiseled into marble the words of Abraham Lincoln: "Let us strive on to finish the work we are in; to bind up the nation's wounds; to care for him who shall have borne the battle, and for his widow, and his orphan—to do all which may achieve and cherish a just and lasting peace, among ourselves, and with all nations."

By now, we know that we will not get such uplifting rhetoric from Donald Trump. For the scientists and many others who have made sacrifices to fight this pandemic, we'd settle for just two words. Two words he has a hard time saying. Two words that he has rarely said to the vaccine scientists who have worked 18-hour days, their kids out of school and family members affected by the pandemic. Or to the health care workers, their faces raw from their protective masks and their souls crushed because they are living in self-isolation to protect their families, who have labored in the wards to bring down the death rate. Two words that he says now, only because he has been diagnosed with COVID-19 himself. Two words he should say constantly to those who have borne the battle of COVID-19. Two words Donald Trump struggles to say: "Thank you."

–H. Holden Thorp*

*Editor's note: An earlier version of this editorial was published online as "Two words Trump can't say" prior to the president's diagnosis of COVID-19.

10.1126/science.abf0577

Published by AAAS

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Di Jiang, Marc S. Lavine (Canada), Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Lucy Hicks, Cathleen O'Grady **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Shanghai), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Joshua Sokol, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Xing Liu **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **WEB DESIGNER** Jennie Pajerowski

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **PRODUCTION SPECIALIST** Kristin Wowk **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer **SALES COORDINATOR** Josh Haverlock

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
sciencemag.org/authors/science-information-authors

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services

science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers

advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org

support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

AAAS BOARD OF DIRECTORS
CHAIR Steven Chu
PRESIDENT Claire M. Fraser
PRESIDENT-ELECT Susan G. Amara
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER
Sudip Parikh
BOARD Cynthia M. Beall
Rosina M. Bierbaum
Ann Bostrom
Stephen P.A. Fodor
S. James Gates, Jr.
Laura H. Greene
Kaye Husbands Fealing
Maria M. Klawe
Robert B. Millard
Alondra Nelson
William D. Provine

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)

Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Institut Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Nenad Ban, *ETH Zürich*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *U. of Texas at Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *VIB*
Kiros T. Berhane, *Columbia U.*
Bradley Bernstein, *Mass. General Hospital*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med. School*
Peer Bork, *EMBL*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC, Irvine*
György Buzsáki, *New York U. School of Med.*
Blanche Capel, *Duke U.*
Annmarie Carlton, *U. of California, Irvine*
Nick Chater, *U. of Warwick*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Carolyn Coyne, *U. of Pittsburgh*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *U. of North Carolina*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Christian Davenport, *U. of Michigan*
Frans de Waal, *Emory U.*
Claude Desplan, *New York U.*
Sandra Díaz, *U. Nacional de Córdoba*
Ulrike Diebold, *TU Wien*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *U. of Penn.*
Jennifer A. Doudna, *UC Berkeley*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Christopher Dye, *U. of Oxford*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Tim Elston, *U. of North Carolina*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *U. of Penn.*
Karen Ersche, *U. of Cambridge*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *The George Washington U.*
Toren Finkel, *U. of Pittsburgh Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *U. of Penn*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Kip Guy, *U. of Kentucky College of Pharmacy*
Taekjip Ha, *Johns Hopkins U.*
Christian Haass, *Ludwig Maximilians U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. College London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Lora Hooper, *UT Southwestern Med. Ctr.*
Fred Hughson, *Princeton U.*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS and U. of Arizona*
Kai Johnsson, *EPFL*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *UC Berkeley*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med. School*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Thomas Langer, *Max Planck Inst. Cologne*
Mitchell A. Lazar, *U. of Penn.*
Wendell Lim, *UC San Francisco*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Washington U. in St. Louis*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marín, *King's College London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth College*
Rodrigo Medellín, *U. Nacional Autónoma de México*
Graham Medley, *LSHTM*
Jane Memmott, *U. of Bristol*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Edward Miguel, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci., Res. & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Harry Orr, *U. of Minnesota*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Margaret Palmer, *U. of Maryland*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Daniel Pauly, *U. of British Columbia*
Samuel Pfaff, *Salk Inst. for Biological Studies*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UC San Francisco*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Félix A. Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial College London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Shimon Sakaguchi, *Osaka U.*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Jürgen Sandkühler, *Med. U. of Vienna*
Erica Ollmann Saphire, *La Jolla Inst.*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Rebecca Sear, *LSHTM*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Washington*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UC San Francisco*
Robert Siliciano, *Johns Hopkins U. School of Med.*
Lucia Sivilotti, *U. College London*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
Mark Smyth, *QIMR Berghofer*
Pam Soltis, *U. of Florida*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Dartmouth College*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Wim van der Putten, *Netherlands Inst. of Ecology*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Kathleen Vohs, *U. of Minnesota*
David Wallach, *Weizmann Inst. of Sci.*
Jane-Ling Wang, *UC Davis* **(S)**
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Jonathan Weissman, *UC San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* **(S)**
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

Published by AAAS

Here today, gone tomorrow

We are destroying the life-support systems of Earth rapidly, making our future uncertain. Ecosystems—the complex sets of organisms that form the globe's living landscape—regulate the atmosphere, water, and soils. They supply humanity its food, most medicines, and many other essential products, and they fill our lives with beauty. But they are falling apart as, one by one, their constituent species are lost. To save what we can and provide our children and grandchildren with a sustainable future, studies must be conducted not only in nature but also, to an increasing extent, on the billions of specimens preserved in the world's natural history collections. For many species barely hanging on in their endangered habitats, these samples will one day be all that we have.

Last week's United Nations Summit on Biodiversity discussed earlier reports that perhaps 1 million of the estimated 8.5 million species of plants, animals, and other organisms are in imminent danger of extinction. Probably as many as half of the populations of organisms that existed half a century ago are already gone. Over the past quarter century, about a quarter of all tropical forests have been lost. Because we have identified no more than a tenth of the estimated tens of thousands of species in those habitats, most that were lost may forever remain unknown.

Unless we control the underlying causes, including overdevelopment and climate change, we are in danger of losing 80% or more of the world's species, the proportion lost 66 million years ago when the dinosaurs became extinct and many of the plants and animals known today began their ascent. We have clearly entered the world's sixth major extinction event.

While we work to forestall this destruction, we must treat the specimens in our collections with ever more care. They are no longer simply samples of wild populations from which more will always be available. As many as half of the specimens likely represent populations that no longer exist, and an increasing number of species as well. What is in our collections will often turn out to be all that remains of organisms that once thrived.

Faced with this grave future, we must find ways to preserve these specimens as well as we can for as long as we can. They are vulnerable to degradation and loss from pests, humidity, fire, and the simple ravages of time. While renewing efforts to protect them, we need to make them more accessible through digitization, including imaging and DNA barcoding (at least a minimal DNA sequence of the representative specimens). We also need to continue targeted sampling, focused on key taxa and habitats.

Some groups of great ecological and environmental importance are dying off too rapidly to ever be completely understood. We have named 25,000 species of nematodes, 64,000 species of mites, and 100,000 species of fungi. Yet each of these groups is estimated to consist of a million or more species, with the number of fungi likely to be 2.2 to 2.6 million. We must sample them and understand them as well as we can before many of their species disappear forever.

> "...we are in danger of losing 80% or more of the world's species..."

To do all of this, we need many more specialists in all regions of the world who are trained in advanced methods for collecting and preserving specimens, and in the exciting new field of "museomics"—determining DNA sequences from old preserved specimens. The standards for the best preservation and training established by the International Society for Biological and Environmental Repositories may help point the way.

Many of the world's biological collections are in institutions that depend in part on attendance for their support. In this time of coronavirus disease 2019 (COVID-19), many of them will fail financially or be unable to continue maintaining their own collections. These and other potentially "orphan" collections have immense value and should be monitored and incorporated as needed into permanent homes. This is likely to be our last chance to know many of Earth's species. We must make the most of it.

–Peter H. Raven and Scott E. Miller

Peter H. Raven
is president emeritus of the Missouri Botanical Garden, St. Louis, MO, USA. peter.raven@mobot.org

Scott E. Miller
is the chief scientist, and interim director of Smithsonian Libraries, Smithsonian Institution, Washington, D.C., USA. millers@si.edu

10.1126/science.abf1185

Published by AAAS

> “Very few don’t have some sort of connection to Big Tech.”
>
> **Doctoral student Mohamed Abdalla**, in *Wired*, about a study he led of faculty members specializing in artificial intelligence at four leading research universities. He found 58% (48 of 83) had received a grant or fellowship from one of 14 large technology companies, which may distort research priorities.

IN BRIEF

Edited by **Jeffrey Brainard**

President Donald Trump removes his mask on a White House balcony after returning from his stay at the Walter Reed National Military Medical Center.

PHOTO: ANNA MONEYMAKER/*THE NEW YORK TIMES*/REDUX PICTURES

POLITICS

Trump’s brief hospitalization for COVID-19 follows bitter debate

The U.S. presidential race was upended in the first days of October as President Donald Trump tested positive for the pandemic virus and spent 3 days in the hospital. He was aggressively treated with two experimental medicines—monoclonal antibodies and the repurposed antiviral remdesivir—and a steroid used in severe COVID-19 cases. Trump returned to the White House on 5 October saying people should not fear the disease. But public health specialists voiced astonishment when he re-entered the building maskless, trailed by questions about his medical condition and a lack of information about how staff members would be protected from infection. All that followed a rancorous first debate on 29 September between Trump and Democratic challenger Joe Biden. Trump mocked Biden for having worn a mask at other times, despite evidence that the precaution reduces transmission of the virus. The president also left scientists puzzled when he described as a “disaster” Biden’s role in the response to the H1N1 swine flu outbreak in 2009. Then-President Barack Obama, whom Biden served under as vice president, declared it a public health emergency 6 weeks before the World Health Organization declared a pandemic. That flu killed an estimated 12,000 Americans—far fewer than the 210,000 U.S. deaths recorded so far from COVID-19.

CDC’s COVID-19 advice faulted

PUBLIC HEALTH | Coronavirus guidelines issued last week by the U.S. Centers for Disease Control and Prevention (CDC) again stirred controversy and concerns that undue political pressure had influenced some of its decisions. CDC announced that on 31 October it will lift an order barring cruise ships from sailing despite a recommendation by its director, Robert Redfield, to extend the ban until February 2021. The industry shut down in March after severe COVID-19 outbreaks occurred on multiple ships. Last week, CDC also drew fire for its updated guidelines on when colleges should test students and faculty and staff members for the pandemic virus. The agency recommended different frequencies of testing, including just a single, initial one, depending on circumstances such as whether students lived in residences with others who tested positive. Critics said the new guidelines should have recommended more regular testing of asymptomatic individuals. CDC addressed another uproar this week by

Published by AAAS

acknowledging evidence that the virus can travel by air and infect people standing more than 2 meters apart in indoor spaces. The agency was faulted last month after it posted, and then withdrew, a draft suggesting otherwise.

U.S. launches zoonosis project

INFECTIOUS DISEASES | An international program to reduce the risk of new zoonotic diseases, allowed to expire by the U.S. government in 2019 but extended until last month, will get a successor. On 30 September, the United States Agency for International Development (USAID) awarded a $100 million grant to help countries in Asia and Africa curb viruses jumping from animals to humans. The 5-year Strategies to Prevent Spillover program will have a different focus from its predecessor, PREDICT, whose termination was criticized by the scientific community: Rather than studying the drivers of spillover, it will seek interventions to reduce viral jumps, a USAID spokesperson says. A key goal is to "help partners at the country level build their expertise and ability to take action," says veterinarian Deborah Kochevar of Tufts University, which leads a 13-institute consortium that won the grant.

Top Soviet-era dissident dies

INTERNATIONAL AFFAIRS | Yuri Orlov, the Russian physicist who championed human rights in the Soviet Union before being exiled in 1985, died on 27 September at age 96. Orlov helped organize the Soviet Union's branch of Amnesty International in 1973 and 3 years later co-founded the Moscow Helsinki Group, which monitored Soviet adherence to the civil rights provisions of the 1975 Helsinki Accords between the Soviet Union and the West. In 1977, Orlov was arrested and sentenced to 12 years of hard labor and exile in Siberia. After coming to the United States in a prisoner exchange, Orlov, an expert in particle accelerators, worked at Cornell University. He didn't think much of Russian President Vladimir Putin, writing in 2004 that "Russia is flying backward in time."

Japan leader blocks appointments

GOVERNANCE | Japan's new prime minister, Yoshihide Suga, has disrupted the process by which scientists are appointed to serve on the governing body of the country's leading academic society, the Science Council of Japan (SCJ). Researchers are criticizing the move as a threat to academic freedom. SCJ makes policy recommendations, promotes scientific literacy and international cooperation, and represents the interests of more than 800,000 scholars in virtually all academic disciplines. The prime minister customarily ratifies appointees recommended by SCJ for its governing body, the General Assembly. But according to an announcement last week, Suga withheld his blessing from six academics, in a list of 105 put forward, who work in the social sciences, law, and the humanities. All six had criticized legislation adopted by Japan's previous government, in which Suga was chief cabinet secretary. His failure to appoint them violated a law governing SCJ, said Satoshi Ihara, secretary general of the Japan Scientists' Association.

THE NOBEL PRIZES

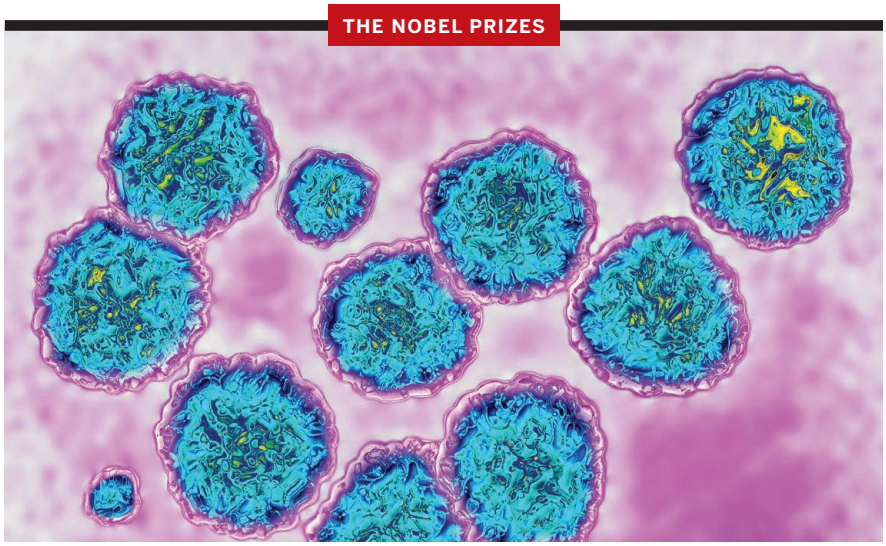


PHOTO: JAMES CAVALLINI/SCIENCE SOURCE

Researchers isolated the hepatitis C virus (shown), leading to a safer blood supply and life-saving treatments.

Medicine prize goes to discoverers of virus that destroys the liver

The Nobel Prize in Physiology or Medicine was awarded this week for the discovery of the hepatitis C virus, one of the most common causes of liver cancer. The prize went to Harvey Alter of the U.S. National Institutes of Health; Michael Houghton of the University of Alberta, Edmonton; and Charles Rice of Rockefeller University.

The hepatitis C virus, transmitted via blood, can cause chronic inflammation of the liver that quietly destroys the organ over decades, ultimately leading to cirrhosis and cancer. The laureates did work over 3 decades to identify the virus and show it was responsible for unexplained cases of hepatitis in people who received blood transfusions. They also developed a test to screen blood donations for the virus, which has nearly eliminated the risk of hepatitis from blood transfusions. Their research ultimately led to a successful treatment for the disease, which has cured millions of people. But about 71 million people worldwide still have chronic hepatitis C, and transmission continues via contaminated medical equipment, sharing drug injection needles, and from infected mothers to newborns during birth. The disease causes few acute symptoms, and testing in many developing countries is limited.

Black hole hunters receive physics prize

The Nobel Prize in Physics has been awarded for pioneering discoveries regarding black holes—self-sustaining gravitational fields so intense that nothing, not even light, can escape.

Roger Penrose, a mathematician at the University of Oxford, received half of the $1.1 million prize for his theoretical work, conducted in part with the late Stephen Hawking, that proved a black hole would be stable and thus could be a real astrophysical object and not a mere mathematical curiosity. Astronomers Reinhard Genzel of the Max Planck Institute for Extraterrestrial Physics and Andrea Ghez of the University of California, Los Angeles, share the other half of the prize for deducing the presence of the supermassive black hole that lies in the heart of our Galaxy. Since the 1990s, Genzel and Ghez have led rival research groups that observed stars there, 26,000 light-years from Earth. They found ones orbiting a heavy, unseen object, called Sagittarius A*, at incredible speeds—some of the most convincing evidence for a behemoth black hole, with the mass of millions of Suns.

Published by AAAS

Mexico axes science funds

POLICY | Mexican scientists this week blasted a move by the national legislature to eliminate 109 trust funds run by public research centers and government institutes, one-third of them devoted to science and technology. The government wants to use the money, some $3 billion in total, for the coronavirus pandemic. The funds support everything from student scholarships and emergency maintenance of equipment to major research projects at dozens of government centers. The money also helps pay for biosecurity and biotechnology research, fighting climate change, and disaster relief. On 6 October, Mexico's Chamber of Deputies approved a bill to terminate the funds, but with "reservations" that require further debate; it is expected to pass in the Senate. The plan is "a brutal blow" and the worst hit to Mexican science in 50 years, says Antonio Lazcano, an evolutionary biologist at the National Autonomous University of Mexico, University City.

Rare poxvirus found in Alaska

VIROLOGY | There's been a new case of infection with Alaskapox virus, a recently discovered pathogen that's related to smallpox. Alaska state health authorities reported on 30 September that they had found the virus in a woman from the Fairbanks area with a mild, gray skin lesion on one arm, similar to one seen in 2015 in the first known patient, also a woman from Fairbanks. Human infections with poxviruses are on the rise, presumably because vaccination against smallpox—which offers some protection against related viruses—was halted after that deadly disease was eradicated 40 years ago. But the Alaska cases are no cause for alarm: There is no evidence the virus can be transmitted between humans—scientists think it came from wild mammals—and the lesions went away by themselves.

Olive ridley sea turtle hatchlings creep seaward just after emerging from eggs on Costa Rica's Pacific Coast.

CONSERVATION

Spy eggs help track turtle poaching

High-tech fake turtle eggs can spy on poachers and wildlife trafficking routes. The real eggs are a delicacy in Central America, and illicit trading of them adds to other hazards to the survival of turtle species that are threatened. Researchers slipped 101 decoy eggs with GPS trackers embedded (left) into nests on four Costa Rican beaches. The scientists tracked five eggs to learn where the poachers took them; the farthest ended up 137 kilometers inland, the multinational team reported on 5 October in *Current Biology*. The researchers did not share this information with authorities, noting ethical concerns; many poachers live in poverty, and in Costa Rica, buying the eggs is not illegal. But, the authors say, the study shows that law enforcement agencies could use the method.

FEATURED INTERVIEW

Fauci: 'Skunk at the picnic'

On 23 September, in the relative calm before President Donald Trump's coronavirus infection was revealed, Anthony Fauci relaxed at home after tangling earlier that day with U.S. Senator Rand Paul (R–KY) during a hearing on COVID-19. Fauci still had 200 emails in his inbox to read that night, but the head of the National Institute of Allergy and Infectious Diseases, who also serves on the White House's Coronavirus Task Force, sat down with *Science* to discuss the pandemic and research on vaccines. (Read the full interview at scim.ag/FauciOctober.) Some excerpts:

On his showdown with Paul: "I said to myself, you know, 'I'm sorry, I'm not gonna disrespect him, but I'm not gonna let him get away with saying things that are cherry-picked data.'" (Paul had suggested that the United States follow Sweden's COVID-19 policies because it had a lower death rate from the disease.)

On speaking bluntly at the White House: "I'm walking a fine line of being someone who is not hesitant to tell the president and the vice president what they may not want to hear. There are some people in the White House, who, even when I first started telling it like it was in the task force meetings, they were like, 'Oh my goodness.' That's when I got that nickname 'the skunk at the picnic.' ... I say, 'I'm sorry, I'm not trying to undermine the president. But there is something that's called reality.'"

On the state of the pandemic: "Yes, there are parts of the country that are doing well. But this country is a big forest, and when you have fires in some parts of the forest, the entire forest is at risk."

PHOTOS: (TOP TO BOTTOM) INGO ARNDT/MINDEN PICTURES; HELEN PHEASLEY

Published by AAAS

In August, raging fires in California destroyed buildings at the Lick Observatory but spared telescopes.

PHOTO: ELINOR GATES/UCO/LICK OBSERVATORY

CLIMATE CHANGE

Astronomy is—and has—a climate problem

Big carbon footprints and rising vulnerabilities push astronomers to take action

By **Daniel Clery**

In August, as wildfires crept close to the historic Lick Observatory near San Jose, California, Claire Max watched as live webcams showed flames edging toward observatory buildings and several telescopes. In the end, firefighters kept the flames at bay. Although two unused buildings were destroyed and several houses were damaged, the working telescopes only had a bit of ash on the mirrors. "We really lucked out," says Max, director of the University of California Observatories, which runs Lick. Coastal California has always experienced cycles of drought and fire, she says. "But it's perfectly plausible for people to say global warming didn't make it any better."

Astronomers have a climate problem. Not only is global warming increasing the frequency of wildfires and the strength of hurricanes that physically threaten observatories, but a changing climate could mar their views by bringing higher temperatures, humidity, and turbulent air closer to their mountaintop perches. Astronomers are also adding to the climate problem themselves, with long flights to remote facilities and meetings and heavy use of energy-hungry supercomputers for cosmic simulations. "We're part of the problem, not of the solution," says Leo Burtscher of Leiden University.

Those concerns were cast in sharp relief by six papers published last month in *Nature Astronomy*. One, on the carbon costs of meetings, emerged directly from the 2019 European Astronomical Society (EAS) meeting in France, which took place during a record-breaking heat wave when temperatures exceeded 45°C. "We were sitting with no air conditioning, sweating through all these interesting talks," Burtscher says. Discussions turned to climate change and the carbon emitted getting everyone to the meeting, and they inspired Burtscher and his colleagues to size up the meeting's travel emissions. They added up to nearly 1900 tons of carbon dioxide (CO_2) equivalent or about 1.5 tons per delegate—roughly the same as emitted by an average resident of India in a whole year.

Other fields of science doubtless generate similar emissions from large meetings, but astronomers' work habits are carbon heavy as well. Another of the six studies found that Australian astronomers each produced 37 tons of CO_2 equivalent per year, of which 60% came from supercomputer usage. "We were surprised how big the supercomputers were. Everyone thought flights would dominate," says lead author Adam Stevens of the University of Western Australia, Perth. A similar analysis in *Nature Astronomy* of the 2018 emissions from Germany's Max Planck Institute for Astronomy (MPIA) showed lower emissions overall (18 tons per astronomer), to which flights were the biggest contributor. The authors say the greater use of renewable energy in Germany may explain some of that difference, but the MPIA astronomers still generated three times the emissions of a typical German.

The climate that astronomers are helping warm is in turn threatening their view of the sky. In another of the papers, MPIA's Faustine Cantalloube and colleagues went through 30 years of weather records from the Paranal Observatory in Chile, operated by the European Southern Observatory (ESO). They found average temperatures there had risen by 1.5°C, more than the 1°C average global rise since preindustrial times. Cantalloube says that's already causing trouble for Paranal's Very Large Telescope—four separate 8.2-meter reflectors. During the day, cooling systems kick in to keep temperatures inside the telescope domes the same as that of exterior air at sunset, to avoid temperature differences that create turbulence when the dome is opened. When daytime temperatures exceed 16°C, the system struggles. "The dome cooling is not good enough," Cantalloube says. "At the time it was built in the 1990s, we were not thinking that [the temperatures] would be so high by 2020."

The hike in temperature has also increased turbulence in the surface layer, the air a few tens of meters above the telescopes.

Published by AAAS

Surface layer turbulence has risen since the 1980s, the researchers found, and although it hasn't yet affected observing, it is worrying, Cantalloube says. "There is clearly something going on very close to the ground," she says.

A hotter climate is expected to raise humidity levels as well, which can lead to increased cloud cover—an obvious problem for astronomers. The water vapor itself can block infrared and microwave radiation, the focus of the Atacama Large Millimeter/submillimeter Array (ALMA), also in Chile. So far, moisture has not increased at Paranal, Cantalloube's analysis shows, or at ALMA, according to a separate study. Climate models are not yet fine-grained enough to predict future moisture trends at the observatories. "We will have to see if climate change will drive more humidity into the region," says ESO atmospheric scientist Angel Otárola.

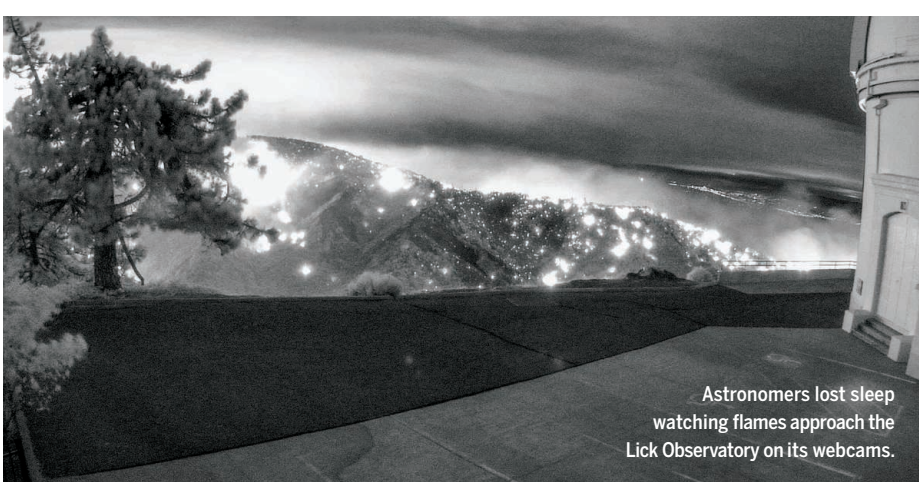
Astronomers are now taking steps to reduce their carbon footprint. In another of the six studies, Simon Portegies Zwart of Leiden University calls for changes in computing strategy. Astronomers should avoid traditional computers and instead use ones that rely on more efficient graphical processor units, Zwart says, although they are harder to program. Astronomers should also abandon popular programming languages such as Python in favor of efficient compiled languages. Languages such as Fortran and C++, Zwart calculates, are more than 100 times more carbon efficient than Python because they require fewer operations. Another option, says MPIA's Knud Jahnke, is to set up supercomputers in Iceland, with its carbon-free geothermal power and cold climate, which reduces cooling needs, or in other countries with plentiful renewable energy.

Major observatories are also taking action. ESO completed a solar power array in 2016 for its La Silla Observatory in Chile and last year inked a deal for a photovoltaic plant to help cool its Extremely Large Telescope, now under construction near Paranal. The Keck and Gemini telescopes in Hawaii and the Murchison Radioastronomy Observatory in Australia have installed solar arrays, too. To cut down on flying, more and more telescopes are making remote observing routine. Max says the University of California built remote observing rooms on its far-flung campuses so researchers didn't need to travel to Lick and other telescopes. Since the arrival of COVID-19, the scheme has been extended to home laptops—"pajama observing," Max calls it. "The days of flying to Hawaii to observe are numbered," she says.

Meetings are another target ripe for reform. "There's a lot of excitement about the potential" of virtual meetings, says Travis Rector of the University of Alaska, Anchorage, who heads the sustainability committee of the American Astronomical Society (AAS). When this year's EAS meeting went virtual because of COVID-19, the team that calculated the carbon costs of the 2019 meeting did a new analysis. Based on a survey of computer and internet usage by delegates and organizers, they calculated emissions of 582 kilograms for the entire meeting, less than one–three-thousandth of the 2019 meeting total. "That really gave a pause for thought," says Mark McCaughrean, an author and a senior adviser at the European Space Agency.

EAS is studying a hybrid format for future meetings, where those farther away would take part virtually. Rector says AAS has contracts that commit it to several years of in-person meetings, but he expects the society to move to virtual meetings after that. Societies, he says, "will learn from each other how to do it better."

This month, Lick workers cleared brush and trees around the site to lessen the risk of future fires. Astronomers need to take action, too, Burtscher says. It's a moral decision—and a practical one, he says. "We need to change in order to continue our professions." ■



Astronomers lost sleep watching flames approach the Lick Observatory on its webcams.

PHOTO: UCO/LICK OBSERVATORY

COVID-19

A call to test new vaccines head to head, in monkeys

Proposed comparison study could clarify safety and which vaccines work best

By **Jon Cohen**

Primate researchers in the United States have banded together in a push for an ambitious monkey study that would do head-to-head comparisons of the leading COVID-19 vaccine candidates. Although 10 candidates are already undergoing large-scale tests in people, proponents of the monkey plan say those clinical trials may not deliver the comprehensive data needed to choose the safest and most effective vaccines. The comparison trial in monkeys, in contrast, could shed light in a matter of weeks on how the candidates stack up on measures including potential side effects, the strength of immune responses they trigger, and how well they protect against infection and disease.

"We should take a cold, hard look at all of the data and ask ourselves, 'What appears to work best?'" says Nancy Haigwood, who directs the Oregon National Primate Research Center and is a key advocate for the comparative monkey study.

The proposed monkey vaccine comparison faces hurdles: It would add to the pressure on the dwindling U.S. supply of research monkeys, potentially delaying research on other diseases, and it does not yet have funding. Haigwood says she expected the U.S. government would gladly support the effort, which would cost an estimated $10 million, compared with the $10 billion the Trump administration's Operation Warp Speed has already devoted to a COVID-19 vaccine push. But facing a lack of interest by current Warp Speed officials, Haigwood and colleagues at the six other national primate research centers are now turning to the National Institutes of Health (NIH) for support.

Most developers of the vaccine candidates in efficacy trials have already published how well each works in monkeys against a "challenge" with SARS-CoV-2—a deliberate

Published by AAAS

A study comparing COVID-19 vaccines would use monkeys at the Oregon National Primate Research Center.

PHOTO: WILLIAM SUTTON/ONPRC/OHSU

exposure to the pandemic coronavirus that causes COVID-19. But the details of how the experiments were conducted and the ways the results were analyzed differ so profoundly that immunologist John Moore of Weill Cornell Medicine says he can't make sense of how the candidates compare. "It's comparing apples to oranges and bananas," says Moore, who has co-authored a review, on preprint.org, that compares the various monkey studies.

The human vaccine trials, for their part, are likely to yield only preliminary signals of efficacy over the next few months, not clear-cut evidence that one or more is safe and protects people. "We're going to get data dribbling in from clinical trials," says Haigwood, a veteran AIDS vaccine researcher.

The data from the many human trials, some in multiple countries, will also be tough to compare. Jay Rappaport, who heads the Tulane National Primate Research Center, notes the trial populations differ and are infected by different variants of SARS-CoV-2. In addition, the human trials—as with the monkey experiments—often use different assays to measure immune responses. "There's so much variation in the primate studies, but there's even more variation in the human studies," Rappaport says.

In contrast to the human trials that must wait for enough participants to become naturally infected to gauge a vaccine's worth, Haigwood says, monkey challenge studies could deliver definitive results quickly. She says the monkey comparison could start as soon as this month and would require only about 6 weeks to vaccinate animals, challenge them, and assess their immune responses and levels of protection.

Yet now that human vaccine efficacy trials are underway, says Moncef Slaoui, Warp Speed's chief scientific adviser, monkey studies won't add much. "Frankly, I did this for 30 years," adds Slaoui, who headed vaccine development at GlaxoSmithKline, "and the primate study is relevant to give you information before the clinical trial. But when you have a phase III trial in tens of thousands of subjects, the relevance of that information is 10 times, 100 times more significant."

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

The backers of the monkey comparison see the situation differently. "I understand that some people say it's too little, too late, but it's not," insists Haigwood, who helped the seven national primate research centers form a consortium to conduct the study. "This is not a competition with Warp Speed. The goal here is to have data that are additive."

The comparison study would likely require a few hundred rhesus macaques, at a time when the long-standing U.S. shortage of monkeys for biomedical research has been exacerbated by China's decision in January to stop exporting all wild animals in the wake of COVID-19. The foreign monkeys went to private labs that do contract research for companies and academic groups, but because those firms have been short on animals, more researchers have turned to the national primate centers for help. Haigwood says the U.S. consortium has many experiments now in the wings that they're willing to delay to conduct the comparative COVID-19 vaccine work. "We have set aside our precious monkeys for this use, because we think it's in the national interest."

The proposed comparative study would use a standardized stock of challenge virus that NIH's National Institute of Allergy and Infectious Diseases (NIAID) recently made, and it would follow consistent protocols to challenge the monkeys and evaluate immune responses to the vaccine candidates. As a first step, Tulane has applied to NIH for a $1.7 million grant to establish a coordinating center to develop standard techniques and data collection for the study. If NIH backs the full study, Haigwood and colleagues expect to test roughly 10 vaccines. Each vaccine challenge study would require eight to 10 rhesus macaques; shared placebo groups would limit the animals needed.

A comparative monkey experiment was part of Warp Speed's early plans developed by Peter Marks of the Food and Drug Administration (FDA). Marks, who left Warp Speed in part because it conflicted with his job as head of the FDA division that oversees vaccines, had wanted to select 14 candidates for phase I and II human studies—early trials that focus on safety and immune responses, not efficacy—and, in parallel, conduct the apples-to-apples monkey studies. After Warp Speed assessed both the monkey and human data, Marks hoped it would pick the best four or five for human efficacy trials. "It may be that at the end of the day, they'll get to a vaccine with 90%, 95% efficacy here, but that [proposed comparative monkey study] was just a way to try to get there," Marks says. "I can't say that I was right or wrong."

NIAID head Anthony Fauci says "it would be worthwhile" to conduct a rigorous comparative study, noting that animal results from vaccine studies for AIDS and other infectious diseases have also been difficult to compare, complicating attempts to translate their results to humans. "It's been one of the banes of our existence."

If NIH won't fund the work, Haigwood says, she'll turn to philanthropic institutions. She also needs to persuade vaccine companies to supply their candidates. "I'll do whatever it takes," Haigwood says, including "twisting the arms" of companies. "We can get a hell of a lot of information before the end of December. Wouldn't you like to have that?" ■

Published by AAAS

The LUX-Zeplin dark matter detector, which will hold 7 tons of liquid xenon, may have no larger successors.

PARTICLE PHYSICS

With to-do list checked off, U.S. physicists ask, 'What's next?'

Giant neutrino experiment is only sure thing for the field

By **Adrian Cho**

As U.S. particle physicists contemplate their future, they find themselves victims of their own surprising success. Seven years ago, the often fractious community hammered out its current research road map and rallied around it. Thanks to that unity—and generous budgets—the Department of Energy (DOE), the field's main U.S. sponsor, has already started on almost every project on the list.

So this week, as U.S. particle physicists start to drum up new ideas for the next decade in a yearlong Snowmass process—named for the Colorado ski resort where such planning exercises once took place—they have no single big project to push for (or against). And in some subfields, the next steps seem far less obvious than they were 10 years ago. "We have to be much more open minded about what particle physics and fundamental physics are," says Young-Kee Kim of the University of Chicago, chair of the American Physical Society's division of particles and fields, which is sponsoring the planning exercise.

Ten years ago, the U.S. particle physics community was in disarray. The high-energy frontier had passed to CERN, the European particle physics laboratory near Geneva, where in 2012 the world's biggest atom smasher, the Large Hadron Collider (LHC), blasted out the long-sought Higgs boson, the last piece in particle physicists' standard model. Some physicists wanted the United States to build a huge experiment to fire elusive particles called neutrinos long distances through Earth to study how they "oscillate"—morph from one of their three types to another—as they zip along. Others wanted the country to help push for the next big collider.

Those tensions came to a head during the last Snowmass effort in 2013, and the subsequent deliberations of the Particle Physics Project Prioritization Panel (P5), which wrote the road map. U.S. researchers agreed to build the neutrino experiment, but make it bigger and better by inviting international partners. They also decided to continue to participate fully in the LHC, and to pursue a variety of smaller projects at home (see table, below). The next collider would have to wait. Most important, DOE officials warned, the squabbling and backstabbing had to stop. In fact, physicists recall, the 2013 process had an informal motto: "Bickering scientists get nothing."

Physicists have just started to build the current plan's centerpiece. The Long-Baseline Neutrino Facility (LBNF) at Fermi National Accelerator Laboratory (Fermilab) in Illinois will shoot the particles through 1300 kilometers of rock to the Deep Underground Neutrino Experiment (DUNE) in South Dakota, a detector filled with 40,000 tons of frigid liquid argon. LBNF/DUNE, which should come online in 2026, aims to be the definitive study of neutrino oscillations and whether they differ between neutrinos and antineutrinos, which could help explain how the universe generated more matter than antimatter.

"The angst in the neutrino community is a lot lower than it was last time," says Kate Scholberg, a neutrino physicist at Duke University. "The DUNE program will be going on at least into the 2030s." However, researchers are already thinking of upgrades to the $2.6 billion experiment, she notes.

In other areas, the future looks less certain. The last time around, for example, scientists had a pretty clear path forward in their search for particles of dark matter—the so-far-unidentified stuff that appears to pervade the galaxies and bind them with its gravity. Researchers had built small underground detectors that searched for the signal of weakly interacting massive particles (WIMPs), the leading dark matter candidate, bumping into atomic nuclei. The obvious plan was to expand the detectors to the ton scale.

Now, two multi-ton WIMP detectors are under construction. But so far WIMPs haven't shown up, and scaling up that technology further "is probably not going to work very well anymore," says Marcelle

Missions accomplished

Major projects prioritized by particle physicists in 2014 are approved, under construction, or up and running.

PROJECT	PURPOSE	STATUS
Long-Baseline Neutrino Facility/Deep Underground Neutrino Experiment	Study how neutrinos change type as they fly from Fermilab in Illinois to South Dakota.	Civil construction begun
High-Luminosity Large Hadron Collider (LHC)	U.S. contributions to upgrades at the LHC	Design work continuing
Camera for Vera C. Rubin Observatory	Survey entire hemisphere of sky every 3 days.	Under construction
Second- and third-generation dark matter detectors	Reach ton-scale detectors for dark-matter particles.	Running or under construction
Next-generation cosmic microwave background experiment	Network of telescopes to study big bang afterglow	Design work continuing
Dark Energy Spectroscopic Instrument	Telescope to study distribution of galaxies and probe space-stretching dark energy	Running
Short-baseline neutrino experiments	Study properties of neutrinos in experiments at Fermilab.	Under construction
Proton Improvement Plan II	New linear accelerator to increase power at Fermilab complex	Design work continuing

CREDITS: (PHOTO) MATTHEW KAPUST/STANFORD UNDERGROUND RESEARCH FACILITY; (DATA) *PARTICLE PHYSICS PROJECT PRIORITIZATION PANEL REPORT* (2014)

Published by AAAS

Soares-Santos, a physicist at the University of Michigan, Ann Arbor. "So we need to think a little bit more out of the box." Researchers are now contemplating a hunt for other types of dark matter particles, using new detectors that exploit quantum mechanical effects to achieve exquisite levels of sensitivity.

A perennial question for the field is what the next great particle collider will be. The obvious need is for one that fires electrons into positrons to crank out copious Higgs bosons and study their properties in detail, says Meenakshi Narain, a physicist at Brown University. But possible designs vary. Physicists in Japan are discussing such a Higgs factory in the form of a 30-kilometer-long linear electron-positron collider. Meanwhile, CERN has begun a study of an 80- to 100-kilometer circular collider. China has plans for a similar circular collider.

However, Vladimir Shiltsev, an accelerator physicist at Fermilab, says those aren't the only potential options. "The real picture is much murkier." Snowmass organizers have received at least 16 different proposals for colliders, including one that would smash together muons—heavier, unstable cousins of electrons—and another that would use photons. Snowmass participants should consider all options, Shiltsev says.

Joe Lykken, Fermilab's deputy director for research, suggests physicists could even push for DOE to support a massive experiment that has nothing to do with particles: a next-generation detector of gravitational waves, spacetime ripples set off when massive objects such as black holes collide. Their discovery in 2015 by the Laser Interferometer Gravitational-Wave Observatory (LIGO) opened a new window on the universe.

LIGO consists of two L-shaped optical instruments with arms 4 kilometers long in Louisiana and Washington; it was built by the National Science Foundation. The next generation of ground-based detectors could be 10 times as big, and might better fit DOE, which specializes in scientific megaprojects, Lykken says. "It starts to sound like the kind of thing that the DOE would be interested in and maybe required for," he says.

During the coming year, Snowmass participants will air the more than 2000 ideas researchers have already proffered in two-page summaries. Then, a new P5 will formulate a new plan. Whatever ideas scientists come up with, to execute their new plan they'll have to maintain the harmony that in recent years has made their planning process an exemplar to other fields. "Being unified is the new norm for us," quips Jim Siegrist, DOE's associate director for high energy physics. "So we have to continue to keep a lid on divisiveness and that'll be a challenge." ■

INFECTIOUS DISEASES

Newly found viruses suggest rubella originated in animals

Two relatives discovered in bats, mice, and zoo animals suggest "German measles" jumped the species barrier

By **Ann Gibbons**

The virus that causes rubella, or German measles, finally has company. Scientists had never identified close relatives of the virus, leaving it as the only member of its genus, *Rubivirus*. But with a report in this week's issue of *Nature*, rubella has gained a family. One of its two newfound relatives infects bats in Uganda; the other killed animals from three different species in a German zoo and was found in wild mice living nearby as well.

The findings strongly suggest that at some point in the past, a similar virus jumped from animals to humans, giving rise to today's rubella virus, the researchers say. Although neither of the new viruses is known to infect humans, the fact that a related virus jumped species raises concerns that the two viruses or other, as-yet-unknown relatives could cause human outbreaks. "We would be remiss not to be concerned, given what's going on in the world today," says epidemiologist Tony Goldberg of the University of Wisconsin, Madison, a senior author of the study.

> **"We would be remiss not to be concerned, given what's going on in the world today."**
> **Tony Goldberg**, University of Wisconsin, Madison

The rubella virus usually causes rashes and fever, but in pregnant women it can lead to miscarriages, stillbirth, and babies born with congenital rubella syndrome, which includes deafness and eye, heart, and brain problems. An estimated 100,000 newborns are affected by the syndrome annually, mostly in Africa, the western Pacific, and the eastern Mediterranean; in many other countries the measles, mumps, and rubella (MMR) vaccine has made it a rarity.

Goldberg and his former graduate student Andrew Bennett discovered one of the new viruses in apparently healthy cyclops leaf-nosed bats, netted at night in Kibale National Park in Uganda. They named it ruhugu virus, after the Ruteete region of Uganda and the local word for bat. The architecture of ruhugu's genome is identical to that of the rubella virus, and 56% of the amino acids in its eight proteins matched those in rubella. The protein that interacts with the host's immune cells was almost identical in both viruses.

As they were getting ready to publish, the researchers learned that a team led by Martin Beer at the Friedrich-Loeffler Institute had detected another rubella relative in brain tissue from a donkey, a kangaroo, and a capybara—a giant rodent native to South America—that all died from encephalitis, an inflammation of the brain, at an unnamed zoo. They found the same virus in wild yellow-necked field mice caught in or near the zoo. The mice appeared to be fine, suggesting they were a natural reservoir from which the virus spilled over to the zoo animals.

Comparing their data, the teams realized their viruses were related, although ruhugu was closer to rubella than the second relative, the rustrela virus, named after a lagoon in the Baltic Sea. The teams decided to publish jointly.

The paper is "really important because there's very little understanding of where rubella came from," says molecular anthropologist Anne Stone of Arizona State University, Tempe. Measles and mumps also came from animals, Goldberg notes. "Now we know that every disease in the letters of the MMR vaccine has a zoonotic origin," he says. Given the genetic distance between rubella and the ruhugu and rustrela viruses, the researchers don't think either of them made the jump to humans—but they suspect they'll find other Rubiviruses if they look closely.

Both viruses bear close watching, researchers say. It's "really interesting" that rustrela was able to infect placental and marsupial mammals, and "was actively jumping between species," says evolutionary virologist Edward Holmes of the University of Sydney. That flexibility could spell trouble, says vaccinologist Gregory Poland of the Mayo Clinic: "Who knows, if it could move from mice to other mammals, could it move to humans?" ■

Published by AAAS

PLANETARY SCIENCE

NASA mission set to sample carbon-rich asteroid

Bennu's thick carbonate veins suggest past life as a water world

By **Paul Voosen**

OSIRIS-REx is ready to get the goods. On 20 October, after several years of patient study of its enigmatic target, NASA's $800 million spacecraft will finally stretch out its robotic arm, swoop to the surface of the near-Earth asteroid Bennu, and sweep up some dust and pebbles. The encounter, 334 million kilometers from Earth, will last about 10 seconds. If it is successful, OSIRIS-REx could steal away with up to 1 kilogram of carbon-rich material from the dawn of the Solar System for return to Earth in 2023.

Since OSIRIS-REx (short for Origins, Spectral Interpretation, Resource Identification, Security, Regolith Explorer) arrived in 2018, Bennu has yielded surprises, not all of them welcome. The 500-meter-wide asteroid was not smooth, as expected, but studded with more than 200 large boulders that could upset the sampling maneuver. And every so often, the asteroid ejected coin-size pebbles, probably propelled by meteoroid impacts or solar heating. The boulder hazard, in particular, forced the team to target an area just 16 meters across for sampling, 10 times smaller than planned. "Bennu has not made things easy for us," says Mike Moreau, the mission's deputy project manager at NASA's Goddard Space Flight Center.

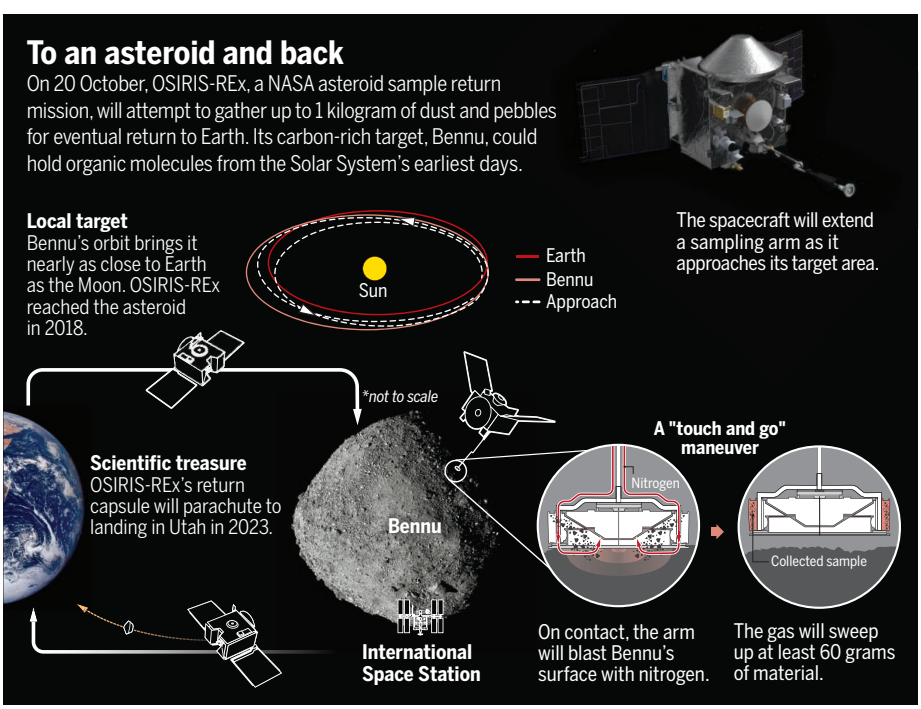
Despite the logistical challenge, the boulders contain a prize: veins of carbonate minerals thicker than your hands, the team reports in one of six studies published today in *Science* and *Science Advances*. The minerals, which precipitate out of hot water, popped out of data gathered during a close flyby of light-colored boulders near the target site, called Nightingale. Researchers believe the veins grew in channels of fluid circulating within Bennu's parent body, a larger planetesimal thought to have formed beyond Jupiter's orbit soon after the dawn of the Solar System 4.56 billion years ago, before being smashed apart in the asteroid belt within the last billion years. Heat from the decay of radioactive elements in its interior presumably drove the churning, and the presence of so much carbonate "suggests large-scale fluid flow, possibly over the entire parent body," says Hannah Kaplan, a planetary scientist at Goddard who led the work.

This ancient water world is consistent with the idea that objects like Bennu delivered much of Earth's water when they struck the planet billions of years ago, says Dante Lauretta, the mission's principal investigator and a planetary scientist at the University of Arizona. The veins also suggest watery bodies like Bennu were a cauldron for the organic chemistry that generated the amino acids and other unusual prebiotic compounds found in carbon-rich meteorites (*Science*, 14 August, p. 760).

OSIRIS-REx won't be sampling the carbonate veins directly: The chamber at the end of its robotic arm is designed to suck up grit smaller than a penny. That's all right, however, because the small pebbles strewn across Nightingale also contain signs of carbonates and other organic molecules, the team reports today. "This gives me a hint that my dream is going to come true," Lauretta says. "I want to bring back something we've never seen before."

The team picked Nightingale for its abundant pebbles and because the site appears young, probably because an impact exposed it in recent geological time, leaving it largely unaltered by bombarding cosmic rays. But navigating the van-size spacecraft to a safe touchdown still won't be easy; the site is ringed with building-size rocks, including one nicknamed Mount Doom, along with smaller boulders throughout. Observations suggest many of these boulders are porous, almost fluffy, and would crumble if touched. But the team doesn't want to take that chance: Using its cameras to navigate, the spacecraft will automatically abort its approach at an altitude of 5 meters if the site appears hazardous.

The entire sampling attempt, lasting 4.5 hours, needs to be autonomous; Bennu is currently five times farther from Earth than Mars, and radio signals take 18 minutes to reach it. After thruster maneuvers bring it to the touch point with Bennu, blasts of nitrogen should push dust and pebbles into the doughnut-shaped collector at the end of the robotic arm. It will be several days before NASA can judge how much was gathered, based on images of the target site and sampling head, and changes in how the spacecraft spins. By the end of the month, managers will decide whether to make a second attempt at a backup site in January 2021. Either way, the spacecraft will leave Bennu next year and head back to Earth. It will arrive in September 2023 and eject the sample capsule, which will parachute to a landing in the Utah desert. ■


To an asteroid and back
On 20 October, OSIRIS-REx, a NASA asteroid sample return mission, will attempt to gather up to 1 kilogram of dust and pebbles for eventual return to Earth. Its carbon-rich target, Bennu, could hold organic molecules from the Solar System's earliest days.
Local target
Bennu's orbit brings it nearly as close to Earth as the Moon. OSIRIS-REx reached the asteroid in 2018.
Sun
Earth
Bennu
Approach
The spacecraft will extend a sampling arm as it approaches its target area.
*not to scale
Scientific treasure
OSIRIS-REx's return capsule will parachute to landing in Utah in 2023.
Bennu
International Space Station
A "touch and go" maneuver
Nitrogen
Collected sample
On contact, the arm will blast Bennu's surface with nitrogen.
The gas will sweep up at least 60 grams of material.

CREDITS: (GRAPHIC) G. GRULLÓN/*SCIENCE*; (IMAGES, LEFT TO RIGHT) NASA; NASA/GODDARD/UNIVERSITY OF ARIZONA (2)

Published by AAAS

Passengers wait for a Stockholm ferry in July. Swedish health officials insist face masks offer a false sense of security and can lead people to forget about social distancing.

PHOTO: JONATHAN NACKSTRAND/AFP VIA GETTY IMAGES

SWEDEN'S GAMBLE

The country's pandemic policies came at a high price—and created painful rifts in its scientific community

By **Gretchen Vogel**

On 5 April, Anders Tegnell, chief epidemiologist for the Swedish public health authority, sent an email to the European Centre for Disease Prevention and Control (ECDC) expressing concern about proposed new advice that face masks worn in public could slow the spread of the pandemic coronavirus. "We would like to warn against the publication of this advice," Tegnell wrote. How much people without symptoms contribute to spread was a "question that remains unanswered," he wrote, and the advice "would also imply that the spread is airborne, which would seriously harm further communication and trust among the population and health care workers."

On 8 April, ECDC published its recommendations anyway, in line with an emerging scientific consensus. Although questions remained, "use of face masks in the community could be considered," it said, "especially when visiting busy, closed spaces." Tegnell still disagrees. "We have looked very carefully. The evidence is weak," he told *Science*. "Countries that have masks are not doing the best right now. It is very dangerous to try to believe that masks are a silver bullet."

Sweden's approach to the coronavirus pandemic is out of step with much of the world. The government never ordered a "shutdown" and kept day care centers and primary schools open. While cities worldwide turned into ghost towns, Swedes could be seen chatting in cafés and working out at the gym. The contrast evoked both admira-

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

Published by AAAS

tion and alarm in other countries, with journalists and experts debating whether the strategy was brilliant—or whether Tegnell, its main architect, had lost the plot.

The country did not ignore the threat entirely. Although stores and restaurants remained open, many Swedes stayed home, at rates similar to their European neighbors, surveys and mobile phone data suggest. And the government did take some strict measures in late March, including bans on gatherings of more than 50 people and on nursing home visits.

Yet Sweden adopted strikingly different policies from those of other European countries, out of a desire to avoid disrupting daily life—and perhaps the hope that, by paying an immediate price in illness, the country could achieve "herd immunity" and put the pandemic behind it.

Swedish authorities actively discouraged people from wearing face masks, which they said would spread panic, are often worn the wrong way, and can provide a false sense of safety. Some doctors who insisted on wearing a mask at work have been reprimanded or even fired.

Until last month, Sweden's official policy stated people without obvious symptoms are very unlikely to spread the virus. So instead of being quarantined or asked to stay home, family members, colleagues, and classmates of confirmed cases had to attend school and show up for work, unless they had symptoms themselves. Testing in Sweden still lags behind many other countries, and in many districts infected people are expected to notify their own contacts—in contrast to, say, Germany and Norway, where small armies of contact tracers help track down people who may have been exposed.

Lena Einhorn (left) is a member of the Vetenskapsforum COVID-19, a group that has harshly criticized policies championed by Sweden's chief epidemiologist, Anders Tegnell (upper right), and his predecessor, Johan Giesecke (lower right).

PHOTOS: (CLOCKWISE FROM LEFT) ASTRID ERIKSSON TROPP/CC 3.0; ANDERS WIKLUND/TT NEWS AGENCY/AFP VIA GETTY IMAGES; IBL/SHUTTERSTOCK

The Swedish approach has its fans. Protesters against coronavirus-related restrictions in Berlin in late August waved Swedish flags. In the United States, a prominent member of President Donald Trump's coronavirus task force, neuroradiologist Scott Atlas, has cited Sweden as a model to follow. The policies also have widespread public support in Sweden, where consensus is prized and criticism of the government is rare.

But within Sweden's scientific and medical community, a debate about the strategy has simmered and frequently boiled over—in the opinion pages of newspapers, within university departments, and among hospital staff. A group of scientists known as "the 22" has called for tougher measures since April, when it published a blistering critique of the country's public health authority, the Folkhälsomyndigheten (FoHM). The group, which has grown to include 50 scientists and another 150 supporting members, now calls itself the Vetenskapsforum COVID-19 (Science Forum COVID-19).

It says the price for Sweden's laissez-faire approach has been too high. The country's cumulative death rate since the beginning of the pandemic rivals that of the United States, with its shambolic response. And the virus took a shocking toll on the most vulnerable. It had free rein in nursing homes, where nearly 1000 people died in a matter of weeks. Stockholm's nursing homes ended up losing 7% of their 14,000 residents to the virus. The vast majority were not taken to hospitals. Although infections waned over the summer, scientists worry a new wave will hit in the fall. Cases are rising rapidly in the greater Stockholm area, where almost one-quarter of the Swedish population lives.

The group's criticism has not been welcomed—indeed, some of the critics say they have been pilloried or reprimanded. "It has been so, so surreal," says Nele Brusselaers, a member of the Vetenskapsforum and a clinical epidemiologist at the prestigious Karolinska Institute (KI). It is strange, she says, to face backlash "even though we are saying just what researchers internationally are saying. It's like it's a different universe."

LENA EINHORN PAID CLOSE attention in January to the news of a new virus spreading in Wuhan, China. Einhorn, who has an M.D./Ph.D. in virology and tumor biology, is better known in Sweden as a filmmaker and book author. "But I can still read a scientific paper," she says. And what she read in *The Lancet* on 31 January was alarming: A model predicted large outbreaks of the new virus in cities around the world. As far as she could see, nothing was being done in Sweden to get ready for the threat.

Concerned, she wrote an email to Tegnell. "I asked, 'Have you seen this paper? Isn't it time we prepare for this?'" Tegnell answered immediately, Einhorn says: "He basically said, 'Well, we shall see. Everyone is trying to apply complex models to very limited data.'" She wrote back emphasizing how easily the virus seemed to spread, including from people without obvious symptoms, and asked about restricting travel from China.

Published by AAAS

Tegnell noted that the World Health Organization (WHO) opposed such measures, she says, then stopped responding. So Einhorn approached Björn Olsen, a professor of infectious diseases at Uppsala University who was raising the alarm in interviews. "What can we do?" she says she asked Olsen.

In late February, during the school holidays, thousands of families went skiing in the Alps—just as reports surfaced about an outbreak in northern Italy. Many had asked whether they should stay home, but health authorities "kept saying, 'No, don't cancel your trip!'" Einhorn says. "It was the middle of that week when the cases in the Italian Alps went boom." As vacationers returned, many asked whether they should quarantine, but FoHM maintained there was no reason to worry.

When 30,000 music fans gathered in a Stockholm arena on 7 March for the national final of the Eurovision Song Contest, "I'm going bonkers," Einhorn says. "I can't sit still." She reached out to a journalist friend and started to write op-eds. Olsen linked her to "a group of desperate scientists," she says. "Suddenly I'm in the middle of an email thread of infectious disease specialists, virologists, epidemiologists," all extremely worried.

On 12 March, as new cases outpaced test capacity, FoHM announced doctors should only test those with severe symptoms, recalls KI immunologist Cecilia Söderberg Nauclér. "I turned to my husband and said, 'They are letting it loose. We are going to crash the health system. We are going to need 500 ICU [intensive care unit] beds and we have 90 in Stockholm.'" On the same day, Norway closed schools, many businesses, and its borders, mirroring measures across Europe.

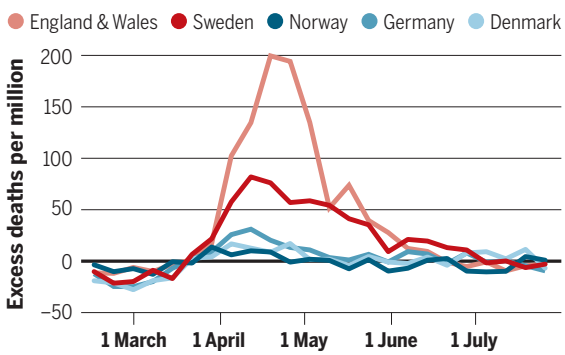
On 15 March, Olsen, Nauclér, and five others warned in an opinion piece in the *Svenska Dagbladet* newspaper that Sweden was just a few weeks behind Italy, where hospitals were already overflowing. Nauclér says she reached Tegnell by phone the next day and told him, "I don't want to argue with you, but you shouldn't be doing what you're doing unless you have data that I don't know about." She says they had a good conversation and Tegnell agreed to a meeting, but it never happened.

The next week, Tegnell announced Sweden would try to "flatten the curve" so the health system would not get overwhelmed with cases. The government limited gatherings to a maximum of 500 people, but day care and schools through ninth grade stayed open. (Upper secondary schools and universities went online.) People should work from home if possible, FoHM said, but tests remained very limited, and close contacts of suspected cases were not asked to stay home unless they had symptoms.

Soon, infections surged. By late March, more than 30 COVID-19 patients were being admitted to ICUs every day. By early April, Sweden was recording about 90 deaths from the virus daily—a significant undercount, critics say, because many died without getting tested. Hospitals did not become as overwhelmed as those in northern Italy or New York City, but that was in part because many severely ill patients weren't hospitalized. A 17 March directive to Stockholm area hospitals stated patients older than 80 or with a body mass index above 40 should not be admitted to intensive care, because they were less likely to recover. Most nursing homes were not equipped to administer oxygen, so many residents instead received morphine to alleviate their suffering. Newspaper reports told stories of people who died after being turned away from emergency rooms because they were deemed too young to suffer serious COVID-19 complications.

Unwelcome distinction

One measure of COVID-19's impact, excess mortality, was far higher in Sweden than in neighboring countries and Germany, but not as high as in England and Wales.



On 25 March, as confirmed cases passed 300 per day, about 2000 scientists signed an open letter calling for stricter control measures. It provoked little reaction. But a scathing op-ed, published by the 22 researchers in the newspaper *Dagens Nyheter* on 14 April, did get noticed. The piece carried the headline "The public health agency has failed. Politicians must intervene." It noted that from 7 to 9 April, more people per million inhabitants had died in Sweden from COVID-19 than in Italy—and 10 times more than in Finland. FoHM officials "have so far not shown any talent for either predicting or limiting" the epidemic, they wrote.

The response was swift. A cascade of columnists and opinion writers criticized the piece for its tone and said the 22 got their numbers wrong. Tegnell said the authors "were not leaders in their field" and claimed they "cherrypicked" days with the highest death tolls. (The scientists replied they had used ECDC statistics and noted there were even more deaths the next week.) The response to the op-ed was "insane," says co-author Jan Lötvall, an allergist at the University of Gothenburg. "A colleague emailed me to say [the article] was shameful, and that we should be loyal and follow the tradition of respecting public health workers."

The frontal attack violated one of Sweden's strongest cultural norms, the taboo on open disagreement, says Andrew Ewing, an analytical chemist at the University of Gothenburg who moved to Sweden from the United States 13 years ago. If a disagreement does arise, "you can never make it personal," says Ewing, who was not part of the original 22 but has since joined the Vetenskapsforum.

"When the debate started, harsh words were exchanged," says Göran Hansson, a cardiac specialist at KI and secretary general of the Royal Swedish Academy of Sciences. But the debate is important, he adds. "Maybe Sweden has too much of a consensus culture. … It's healthy for science to have discussions. One thing we don't need in this situation is silencing of views, especially from those with expertise."

Healthy or not, Brusselaers says she also faced backlash from colleagues and was publicly reprimanded by her department chair for being a "troublemaker" and "a danger to society." "A colleague told me, 'We have to stick with [FoHM] and defend it,'" she says. The situation prompted her to return to her native Belgium, where she now has a position at the University of Antwerp, although she is also keeping her group at KI. "I just didn't expect this reaction in Sweden," she says. "I never felt like such a foreigner as I did over the past few months."

Those who challenged the recommendations against face masks faced a similar backlash. Agnieszka Howoruszko, an ophthalmologist at a regional hospital in Landskrona, began to wear a mask in March when she examined patients. "My manager reprimanded me twice," she says. Howoruszko held her ground. "I said, 'I'm sorry, if I can't wear it, I cannot work. Many of my patients are elderly and in high-risk groups." The manager relented and allowed the clinic's doctors (but not other staff) to wear masks. "We are the only eye clinic in our province" to take that step, she says.

Dorota Szlosowska, a pulmonologist who had been working at Sundsvall regional hospital, shared an email with *Science* stating

CREDITS: (GRAPHIC) X. LIU/*SCIENCE*; (DATA) THE HUMAN MORTALITY DATABASE. U.K. OFFICE FOR NATIONAL STATISTICS

Published by AAAS

that one of the reasons her contract wasn't renewed was that "she walked around with a mask," which the email said made her look unfriendly and made it hard for patients to understand her. Björn Lindström, an ophthalmologist at Falu lasarett, a hospital in central Sweden, says he is the only one in his clinic who wears a mask. In a letter in *Dagens Nyheter*, Lindström has argued that the failure of health care workers to adopt masks violates Sweden's patient safety act, meant to prevent patients from being harmed while receiving care.

HARM SEEMS to have occurred. The Falu lasarett announced last week that it had been fighting a COVID-19 outbreak in its cardiac ward for 3 weeks, with 10 patients and 12 staff infected so far. As of 27 September, staff will "use protective visors when working closely with patients," the hospital said. The Swedish Health and Social Care Inspectorate told *Science* it is investigating 17 hospital and clinic-based outbreaks. In September, Ryhov community hospital in Jönköping announced 20 patients and 40 staff members had been infected in an outbreak in the hospital's orthopedic ward in May. Five patients died and one is still hospitalized. (The hospital said it had followed FoHM's policies.) At least three patients have reportedly died of COVID-19 after being infected at the university hospital in Lund.

FoHM's decision to keep schools open despite surging cases may also have added to the spread. A report from the agency itself, released in July, compared Sweden with Finland, which closed its schools between March and May, and concluded that "closing of schools had no measurable effect on the number of cases of COVID-19 in children." But few Swedish children were tested in that period, even if they had COVID-19 symptoms. When new FoHM guidelines allowed symptomatic children to be tested in June, cases in children shot up—from fewer than 20 per week in late May to more than 100 in the second week of June. (FoHM reversed course in July and returned to recommending that children under 16 not be tested.)

Indirect data suggest children in Sweden were infected far more often than their Finnish counterparts. The FoHM report says 14 Swedish kids were admitted to intensive care with COVID-19, versus one in Finland, which has roughly half as many schoolchildren. In Sweden, at least 70 children have been diagnosed with multisystem inflammatory syndrome, a rare complication of COVID-19, versus fewer than five in Finland.

In the population as a whole, the impact of Sweden's approach is unmistakable. More than 94,000 people have so far been diagnosed with COVID-19, and at least 5895 have died. The country has seen roughly 590 deaths per million—on par with 591 per million in the United States and 600 in Italy, but many times the 50 per million in Norway, 108 in Denmark, and 113 in Germany.

Another way to measure the pandemic's impact is to look at "excess deaths," the difference between the number of people who died this year and average deaths in earlier years. Those curves show Sweden did not suffer as many excess deaths as England and Wales—whose tolls were among Europe's highest—but many more than Germany and its Nordic neighbors (see graphic, p. 161). Immigrant communities were hit very hard. Between March and September, 111 people from Somalia and 247 from Syria died, compared with 5-year averages of 34 and 93, respectively.

TEGNELL HAS SAID repeatedly that the Swedish strategy takes a holistic view of public health, aiming to balance the risk of the virus with the damage from countermeasures like closed schools. The goal was to protect the elderly and other high-risk groups while slowing viral spread enough to avoid hospitals being overwhelmed. Protecting the economy was not the aim, he says. (Initial data suggest Sweden's economy contracted about as much as its immediate neighbors' as exports and consumer spending dropped.)

Sweden's light approach is more sustainable than the harsher methods used in other countries, Tegnell also argues. He regrets the death toll in nursing homes, he told *Science*, and says Sweden should have made it easier financially for caregivers to stay home. "It was a very bad situation for a month," he says, "but after that it changed completely." Once strong restrictions were in place, transmission in nursing homes "became lower than in the community." Tegnell has also said he suspects the number of infections and deaths in other countries will eventually match Sweden's. Einhorn finds this absurd: "If Norway ever catches up to Sweden in the proportion of people killed by COVID-19," she says, "I'll eat my hat."

Many of Tegnell's critics say FoHM had an unspoken agenda: to reach herd immunity. Sweden wouldn't be the only country to consider that strategy: British Prime Minister Boris Johnson toyed with the idea before rejecting it (and contracting COVID-19 himself). Dutch Prime Minister Mark Rutte explicitly said achieving herd immunity would help protect the economy before also abandoning the idea.

Herd immunity is still not well understood, but scientists estimate that in the case of COVID-19, between 40% and 70% of a population would have to be immune to arrest the spread of the virus. Many scientists say reaching that percentage without the help of a vaccine would cause far too many deaths and long-term side effects.

PHOTO: DAVID LIDSTROM/GETTY IMAGES

Tegnell has consistently denied that herd immunity is his goal. But emails released in late July after journalists requested them under open records laws show he discussed the idea. In an exchange on 14 and 15 March with the head of Finland's public health agency, Tegnell speculated that "one point would be to keep schools open to reach herd immunity faster." When the Finnish colleague said models suggested closing schools would decrease infection rates among the elderly by 10%, Tegnell replied: "Ten percent might be worth it?" (Tegnell says he was only speculating, and the prospect of reaching herd immunity was irrelevant to the decision to keep schools open.)

Tegnell's thinking appears to have been shaped by his predecessor, Johan Giesecke, an epidemiologist and professor emeritus at KI with whom he exchanged many emails. Giesecke has been a vocal defender of FoHM's strategy, which he praised in a 5 May article in *The Lancet*. He said the vi-

Published by AAAS

A senior couple dines separately from the rest of their family on 18 April in Östersund, Sweden. The Swedish government never ordered a lockdown, but many citizens reduced their contacts anyway, data suggest.

rus was "an invisible pandemic" in which 98% to 99% of infected people don't realize they have been infected. "Our most important task is not to stop spread, which is all but futile, but to concentrate on giving the unfortunate victims optimal care," he wrote. (Giesecke stated he did not have any conflicts of interest, but his correspondence with Tegnell revealed he had been a paid consultant for FoHM since March. Giesecke told *Science* he sees no conflict.)

Giesecke, a member of WHO's Strategic and Technical Advisory Group for Infectious Hazards, is still advising a similar approach to governments elsewhere. On 23 September, he told an Irish parliamentary committee that Ireland should aim for "controlled spread" in people under age 60 and "tolerable spread" among those over age 60, though in a later interview he backed off, saying Ireland had to decide policies for itself.

Giesecke and Tegnell believed herd immunity would arrive quickly. In the *Lancet* article, Giesecke claimed about 21% of residents of Stockholm county had already been infected by the end of April; Tegnell predicted 40% of them would have antibodies by the end of May. When initial studies showed the number was actually about 6% in late May, Tegnell said immunity was hard to measure. FoHM continued to say Swedes had built up immunity, but in September it backtracked, estimating that "just under 12%" of Stockholm residents, and 6% to 8% of the Swedish population as a whole, had antibodies to the virus by mid-June.

If herd immunity is beginning to kick in, it should become visible in Sweden's case numbers. Cases fell from a record 1698 on 24 June to about 200 per day in early September, and the percentage of positive tests reached a record low of 1.2%. Some speculate that Sweden's summer traditions may have helped: Hundreds of thousands leave cities and towns for remote cabins in what amounts to 3 months of national social distancing.

At the time, numbers elsewhere in Europe were beginning to soar again, especially among young adults, whereas those in Sweden remained stable. But over the past few weeks, infections in Sweden have started to rise as well. On 25 September, FoHM reported 633 new cases nationwide in 1 day. Stockholm's rates have nearly tripled in 2 weeks, from 334 in the second week of September to 967 last week. Whether immunity is making a big difference remains to be seen.

THE SWEDISH EXPERIMENT is coming to an end, as its policies fall in line with those of its neighbors. FoHM officials are "quietly changing their approach," Einhorn says. The country has boosted test rates; at roughly two tests per 1000 inhabitants per day, Sweden's testing rate is almost on par with Norway's—although it is only one-quarter of Denmark's. The recommendation against testing children between ages 6 and 16 was lifted for a second time in September. (FoHM says this is so children with mild symptoms can return to school more quickly if their test is negative.) Children under 6 are still not tested unless severely ill.

The drop in cases allows Sweden to start to use its contact tracing system, in place for other diseases, for COVID-19, Tegnell says: "Before, we just didn't have the capacity." And on 1 October, FoHM announced family members of confirmed cases should stay home for 7 days, even if they don't have any symptoms—although children through ninth grade should still go to school.

FoHM should go much further, Hansson and a colleague said in an August opinion piece, for instance by limiting public transport to 50% capacity, recommending masks, and asking travelers from hard-hit regions abroad and all contacts of known cases to quarantine. On 25 September, Hansson announced the Royal Swedish Academy had assembled an expert group to compare the Swedish response with that of other countries and recommend how "researchers can best contribute to future crisis situations."

FoHM "should have listened more carefully to the scientific community both inside and outside the country," Hansson says. Still, he predicts the rifts will eventually heal. "I am sure we will continue to argue, but I don't see permanent damage," he says. "We'll move on. We'll go back to complaining about grants."

But Ewing worries the fight has left permanent scars. He says at least three more members of the Vetenskapsforum are considering leaving Sweden, as Brusselaers did. And even if it turns out that the country has built up enough immunity to evade a new wave of disease, he says, the price has been too high. "I worry that countries around the world are going to say, 'We can try what Sweden did.' But we have killed too many people already." ■

Published by AAAS

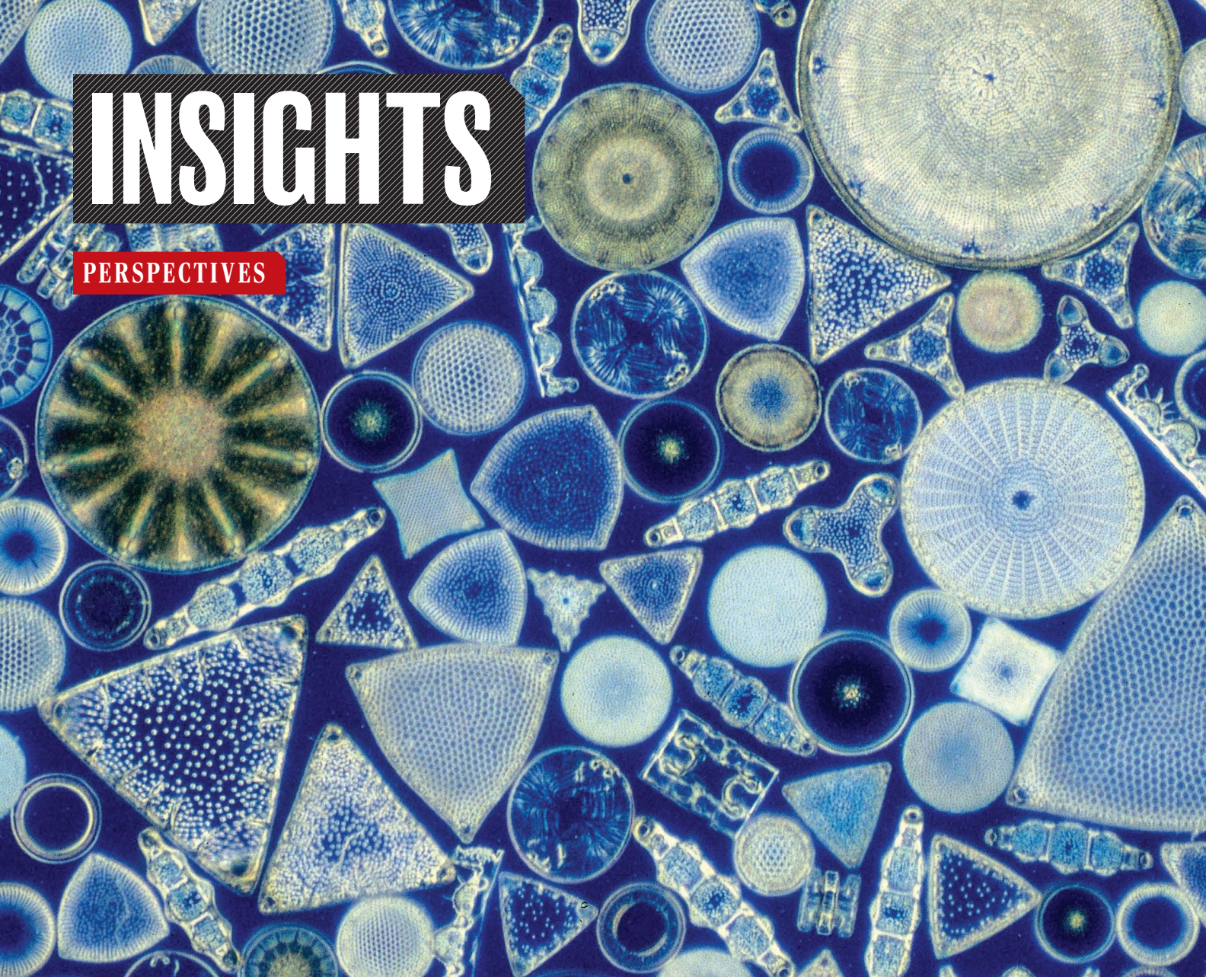


PHOTO: M. I. WALKER/SCIENCE SOURCE

COMMUNITY ECOLOGY

Novel communities are a risky business

High turnover leads to novel combinations of species and involves high extinction

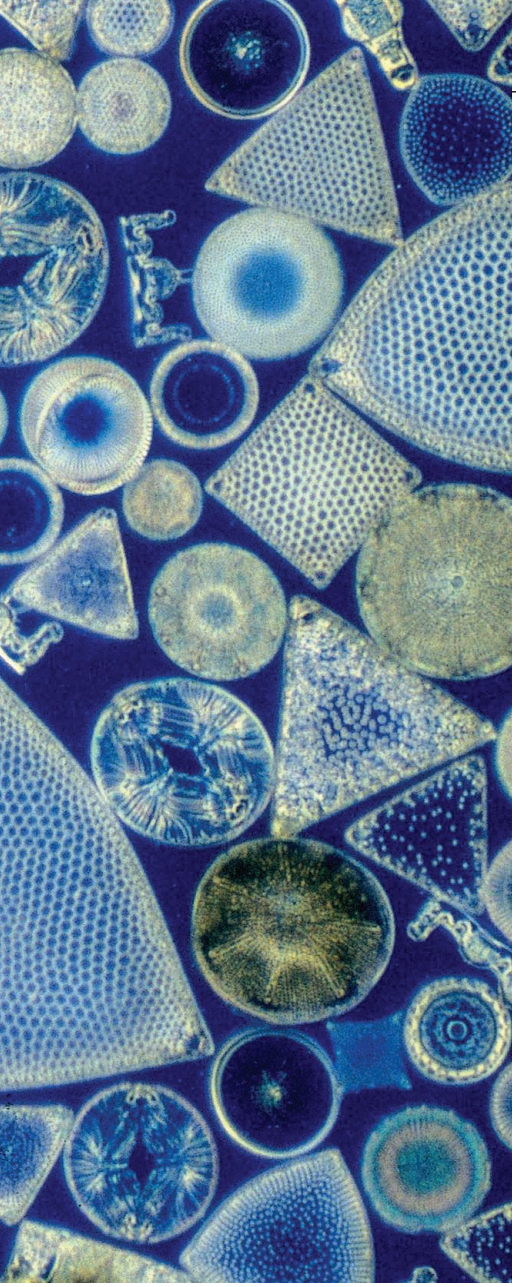
By **Maria Dornelas**[1] *and* **Joshua S. Madin**[2]

As in business, most action in biodiversity happens in turnover. Net changes in balance give us some indication of performance. However, knowing the volumes of income and payments that make up that balance conveys much more information about how a business is operating. The same is true of ecological communities: Knowing net changes in the number of species is useful, but the gains and losses of species encapsulated in turnover hold essential information about biodiversity change. High turnover is emerging as the signature of biodiversity patterns in the Anthropocene (*1*–*3*), raising questions as to the causes and consequences of high turnover in biodiversity. On page 220 of this issue, Pandolfi *et al.* (*4*) look into the deeper past to identify when turnover is so increased that it results in completely new combinations of species, known as novel communities.

The stability of ecological communities is a controversial topic. Under normal conditions, there is some degree of turnover; environmental variation and chance events lead to some fluctuation in which species exist in any particular community. One view is that this variation occurs within bounds because ecological communities exist in dynamic equilibrium in configurations called stable states. Extreme disturbances and persistent forces can lead to abrupt changes in com-

[1]Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK. [2]Hawai'i Institute of Marine Biology, University of Hawai'i at Manoa, Honolulu, HI, USA. Email: maadd@st-andrews.ac.uk

Published by AAAS



Examination of global marine plankton communities over the past 65 million years (such as fossil diatom communities, shown) suggests that change in community composition leads to further change, including extinction.

munities called phase shifts (*5*), which are characterized by very high turnover in composition. Modern examples of this phenomenon are shifts between coral- and algal-dominated systems in tropical reefs or between clear and turbid lakes (*5*). A contrasting view is that community composition changes continuously, determined by a complex array of continuously changing conditions interacting with demographic variation of each species (*6*). This view is supported by growing empirical evidence that discrete community stable states, and shifts between them, are rare or nonexistent (*7*). Even gut microbiomes, originally thought to be relatively stable, show extraordinary turnover (*8*), with communities measured weeks apart bearing very little resemblance to one another (*9*). Looking at how communities have changed in the past can help us understand the mechanics of stability.

The rates of change in species losses and gains that underpin turnover have varied greatly in the past (*10*). The biggest turnover events occurred during mass extinction events (*11*). However, in their study of the past composition of marine plankton communities, Pandolfi *et al.* identify several transitions in the past 65 million years when increased turnover led to the appearance of novel communities. These transitions are rare in the time period examined but highly influential in that they disrupt the communities of species that are found together. For example, the biotic interactions that are necessary for ecosystems to function are disrupted during periods of increased turnover (*12*). Critically, Pandolfi *et al.* find evidence that novelty generates more novelty, with transitions among novel states being 2 to 10 times more likely than one would expect by chance. The transience of novel communities is perhaps indicative of restructuring biotic interactions. Using the quantitative approach proposed by Pandolfi *et al.*, it will be possible to examine whether periods of transition to novel communities coincide with periods of disruption in ecosystem functions.

Transitions to novel communities are underpinned by enhanced rates of local extinctions and colonizations (see the figure). Four main types of process have been proposed to mediate community dynamics: drift, selection, dispersal, and speciation (*13*). Drift entails random changes in species abundances; selection refers to differences among organisms that determine their ability to leave descendants; dispersal relates to movement of organisms across different communities; and speciation is the process of generating new species. Drift and selection control extinction rates, and colonization can arise either from dispersal of existing species or speciation. All four processes are likely to accelerate in transitions to novel communities. For example, 10 million to 12 million years ago, climatic change unleashed selection forces, in combination with enhanced dispersal facilitated by new land bridges, which led to novel communities in North America (*12*). Pandolfi *et al.* find that transitions to novel communities involve the highest rates of local extinctions and originations (relative to emigrations and immigrations).

Evidence is clear that the composition of ecological communities is not static. The history of life on this planet is full of examples of extraordinary communities that no longer exist. Benthic marine faunas of the Cambrian bear little resemblance to those of the modern because of the extinction or major decline of groups such as the trilobites and brachiopods and the radiation of groups such as snails, clams, and corals (*11*). Pandolfi *et al.* add to growing evidence that turnover occurs even during relatively stable periods of the past. This constant change in composition has important implications for conservation goals because it challenges the idea of baseline communities with any particular composition. Communities have always changed in their composition and should continue to do so. A nostalgic longing for a lost Garden of Eden, which permeated the roots of the conservation movement, is not supported by what we know of the past and expect in the future.

Understanding past transitions to novel communities has important implications for current trends in biodiversity. Cases of complete turnover in marine assemblages (*3*), combined with evidence of accelerating extinction and colonization rates (*14*), add to the evidence that many communities are currently undergoing similar transitions. Lessons from the past warn us that more change is likely to unfold. The past also tells us that although biodiversity is always changing, not all change is the same. The dynamic nature of community composition provides an argument for conservation goals to focus on rates of change (*15*) rather than aiming for particular states and for biodiversity science to scrutinize the mediating processes (*13*) responsible for changing rates. ■

REFERENCES AND NOTES

1. M. Dornelas *et al.*, *Science* **344**, 296 (2014).
2. H. Hillebrand *et al.*, *J. Appl. Ecol.* **55**, 169 (2018).
3. S. A. Blowes *et al.*, *Science* **366**, 339 (2019).
4. J. M. Pandolfi *et al.*, *Science* **370**, 220 (2020).
5. M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, *Nature* **413**, 591 (2001).
6. D. E. Schindler, J. B. Armstrong, T. E. Reed, *Front. Ecol. Environ.* **13**, 257 (2015).
7. H. Hillebrand *et al.*, *Nat. Ecol. Evol.* 10.1038/s41559-020-1256-9 (2020).
8. S. Priya, R. Blekhman, *Genome Biol.* **20**, 150 (2019).
9. J. R. Johnson *et al.*, *J. Clin. Microbiol.* **46**, 4078 (2008).
10. J. Alroy *et al.*, *Science* **321**, 97 (2008).
11. J. Sepkoski Jr., R. K. Bambach, D. M. Raup, J. W. Valentine, *Nature* **293**, 435 (1981).
12. J. L. Blois, P. L. Zarnetske, M. C. Fitzpatrick, S. Finnegan, *Science* **341**, 499 (2013).
13. M. Vellend, *Q. Rev. Biol.* **85**, 183 (2010).
14. M. Dornelas *et al.*, *Ecol. Lett.* **22**, 847 (2019).
15. D. Leclère *et al.*, *Nature* 10.1038/s41586-020-2705-y (2020).

ACKNOWLEDGMENTS

M.D. and J.S.M. are supported by a National Science Foundation–Natural Environment Research Council Biological Oceanography grant (1948946). M.D. is supported by Leverhulme Trust Research Centre–the Leverhulme Centre for Anthropocene Biodiversity and a Leverhulme Research Grant (RPG-2019-402).

10.1126/science.abe4727

Published by AAAS

INFECTIOUS DISEASE

Making a bed for viral infections

An inflamed mucosal niche is permissive for symptomatic respiratory syncytial virus infection

By **Ananda S. Mirchandani** *and* **Sarah R. Walmsley**

In the past 100 years, humans have harnessed the power of antibiotics to save countless lives from bacterial infection, however, our capacity to treat viral illnesses lags far behind. Prevention through vaccination has been the most effective method to tackle viral infections, yet vaccines have not always been achievable, as in the case of respiratory syncytial virus (RSV) (*1*). In most people, RSV infection causes a mild "cold"-like illness, and by age 2, most children will have been infected by RSV (*2*). Although most develop only mild symptoms, some children will develop severe lung inflammation (bronchiolitis) that can be fatal. Therefore, determining the risk of severe illness would transform the management of RSV. On page 188 of this issue, Habibi *et al.* (*3*) identified an inflammatory signature in adults susceptible to symptomatic RSV infection, which they ascribe to activated neutrophils. Their findings suggest that the immune environment in which the virus is received may dictate the outcome.

Neutrophils are often overlooked as mediators of the immune response to viruses, even though they are the predominant immune cell in the lungs of individuals with RSV-induced bronchiolitis (*4*) and despite measurements of neutrophil-associated mediators in the lung correlating with RSV disease severity (*5*). Neutrophils have been shown to exacerbate epithelial cell damage but reduce viral titers in RSV infection in vitro, demonstrating how their presence can alter the course of the infection once it has commenced (*6*). Furthermore, neutrophils express most Toll-like receptors (*7*), which recognize pathogens, and they have been shown to respond to other single-stranded RNA viruses, such as influenza A virus (*8*). Thus, it is likely that neutrophils can play an active part in immune responses to other viral infections, including RSV.

Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK. Email: sarah.walmsley@ed.ac.uk

The timing of neutrophil presence in tissues, as well as their activation status, appears to be key. It has been demonstrated that depletion of nonactivated neutrophils before infection and thereafter, or enhancement after infection, does not worsen RSV outcomes in mice (*9*). Habibi *et al.* inoculated 58 healthy adults with RSV and assessed the preinfection transcriptional landscape of their nasal mucosa, identifying a specific inflammatory milieu in those who developed symptomatic RSV infection. In mouse models, preinfection expansion of lung neutrophil numbers with C-X-C motif chemokine ligand 1 (CXCL1) led to increased RSV infection–associated weight loss, which was partially rescued by blocking neutrophil recruitment to the lung. This suggests that an aberrant inflammatory milieu created by activated neutrophils is necessary to drive symptomatic RSV infection. This is supported by the finding of higher preinoculation levels of the neutrophil-associated proteases lipocalin-2 (LCN2) and myeloperoxidase (MPO) in the nasal mucosa of the individuals who developed RSV infection. Moreover, CXCL1 enhanced MPO production in mice, beyond the levels seen in RSV infection alone, suggesting that CXCL1 per se was altering the lung niche.

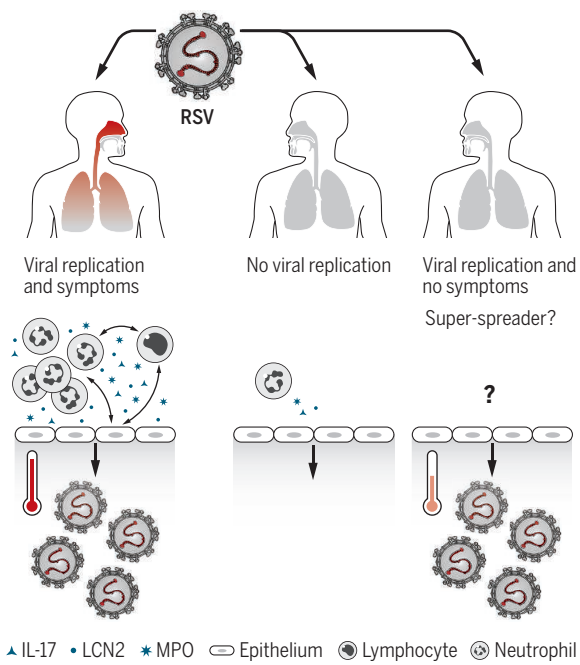
While the timing of the inflammatory milieu appears to be important in allowing the establishment of a symptomatic RSV infection, the dynamic changes in the immune responses recorded in the cohort by Habibi *et al.* are also insightful. Those who developed RSV infection had lower intranasal levels of the archetypal "cytokine storm" cytokines, interleukin-1β (IL-1β), IL-10, and IL-6, in the presymptomatic phase of the infection. This finding emphasizes that immune responses are dynamic and that snapshots in disease processes may be misleading, especially when cross-talk between innate and adaptive immune responses is being ascertained. Furthermore, although an increased preinfection IL-17 level in nasal mucosa was associated with RSV infection, this cytokine was diminished during the incubation period in these people, again underpinning the importance of the timing of these responses in dictating outcomes.

The role of IL-17 as an indirect neutrophil chemoattractant is established, however, the exact source of this cytokine in the nasal mucosa was not determined by Habibi *et al.* Both innate (e.g., γδ T cells) and adaptive (e.g., $CD4^+$ T cells) lymphocytes have been shown to be the predominant producers of IL-17 (*10*), and airway epithelial cells express the IL-17 receptor. A role for IL-17 in immune responses to bacterial infection has been shown (*11*). Yet, whether IL-17 (unlike its close relative IL-22) has a role in epithelial homeostasis remains unclear. Further studies to determine whether IL-17 plays an active role in mucosal homeostasis and in creating an RSV-permissive milieu is warranted, especially as it could allow patient risk stratification.

The capacity to determine the susceptibility to viral infection on an individual level has never been so attractive, given the global impact of the severe acute respiratory

Mucosal niches

The presence of inflammatory mediators such as the neutrophil chemokine interleukin-17 (IL-17) and the neutrophil-associated proteases lipocalin-2 (LCN2) and myeloperoxidase (MPO) in the nasal mucosa of adults is associated with symptomatic respiratory syncytial virus (RSV) infection. Dissecting the mechanisms and immune responses that also allow asymptomatic infection and viral shedding is an important future aim.



GRAPHIC: A. KITTERMAN/*SCIENCE*

Published by AAAS

syndrome coronavirus 2 (SARS-CoV-2) pandemic. Undoubtedly, a number of similarities can be drawn between the viral illnesses induced by RSV and SARS-CoV-2. Both of these enucleated RNA viruses cause a more severe infection in men, patients with underlying immunosuppression, and those with chronic cardiovascular disease (*12, 13*). However, the age distribution of the severe spectrum of disease is somewhat different. Although both viruses can severely affect older people, SARS-CoV-2 does not appear to be as pathogenic in younger children, albeit infants are still at risk (*14, 15*). Whether the findings of Habibi *et al.* therefore apply to other respiratory viruses, or indeed other age groups, remains to be elucidated.

Much of our understanding of immune responses to pathogens, including viral pathogens, has been dissected during infection stages. From a population perspective, however, prevention and risk stratification remain the cornerstone for improving outcomes, whether from established and endemic pathogens such as influenza viruses and RSV, or from emerging pathogens such as SARS-CoV-2. This study also identified a subgroup of people who developed asymptomatic RSV replication. Dissecting the mechanisms and immune responses that also allow this outcome would greatly enhance our capacity to control the spread of viral pathogens on a global scale.

The findings of Habibi *et al.* provide a strong start to further our understanding of how the immune landscape before an insult is just as important in shaping outcomes as the responses to the pathogen itself (see the figure). It is a reminder too, of the dynamic complexity of the immune system, and of the work that lies ahead. ■

REFERENCES AND NOTES

1. P. L. Acosta, M. T. Caballero, F. P. Polack, *Clin. Vaccine Immunol.* **23**, 189 (2015).
2. W. P. Glezen, L. H. Taber, A. L. Frank, J. A. Kasel, *Am. J. Dis. Child.* **140**, 543 (1986).
3. M. S. Habibi *et al.*, *Science* **370**, eaba9301 (2020).
4. M. L. Everard *et al.*, *Arch. Dis. Child.* **71**, 428 (1994).
5. K. Yasui *et al.*, *Pediatr. Int.* **47**, 190 (2005).
6. Y. Deng *et al.*, *J. Virol.* **94**, e02161 (2020).
7. F. Hayashi, T. K. Means, A. D. Luster, *Blood* **102**, 2660 (2003).
8. J. P. Wang *et al.*, *Blood* **112**, 2028 (2008).
9. F. Kirsebom, C. Michalaki, M. Agueda-Oyarzabal, C. Johansson, *Sci. Rep.* **10**, 1110 (2020).
10. S. J. Gurczynski, B. B. Moore, *Am. J. Physiol. Lung Cell. Mol. Physiol.* **314**, L6 (2018).
11. K. Chen *et al.*, *Cell Host Microbe* **20**, 596 (2016).
12. E. J. Williamson *et al.*, *Nature* **584**, 430 (2020).
13. A. R. Falsey, P. A. Hennessey, M. A. Formica, C. Cox, E. E. Walsh, *N. Engl. J. Med.* **352**, 1749 (2005).
14. O. V. Swann *et al.*, *BMJ* **370**, m3249 (2020).
15. T. Shi *et al.*, *Lancet* **390**, 946 (2017).

ACKNOWLEDGMENTS

S.R.W. holds a Wellcome Trust Senior Clinical Fellowship (209220). A.S.M. holds a Wellcome Trust Postdoctoral Training Clinical Fellowship (110086).

10.1126/science.abe3685

INFECTIOUS DISEASE

Inhibiting Ebola virus and SARS-CoV-2 entry

Screening identifies CD74 as a cellular antiviral protein against Ebola virus and SARS-CoV-2

By **Alexandra I. Wells**[1,2] *and* **Carolyn B. Coyne**[1,2]

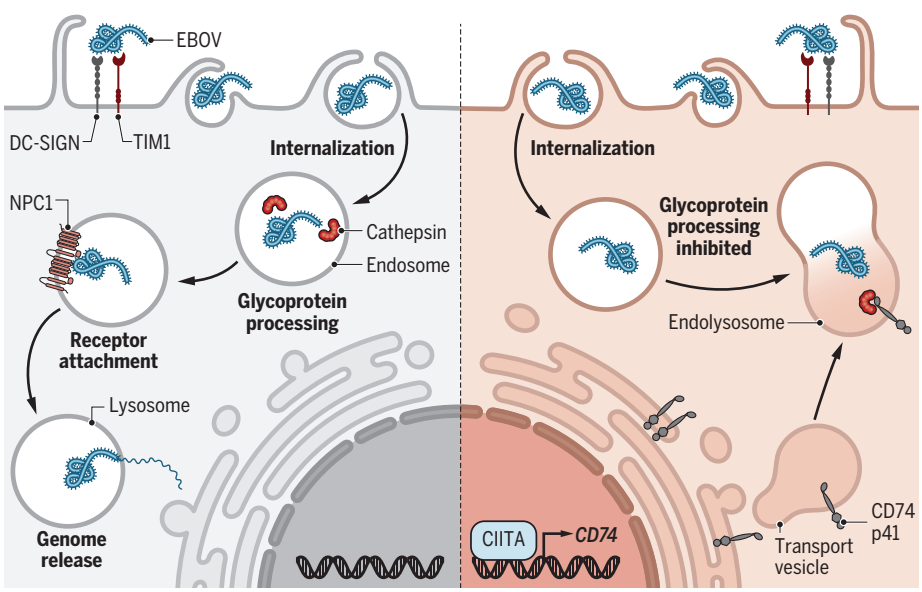
The mechanisms by which cells defend against many viruses remain largely unknown. Defining these mechanisms is important not only for understanding viral pathogenesis but also for informing the development of antiviral therapeutics. The concerted efforts of antiviral factors within cells are central to host cell defense. Without these factors, the cell remains defenseless against potentially harmful pathogens. Understanding how the cell defends itself is particularly important for viruses that have the potential to affect global health, such as Ebola virus (EBOV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). On page 241 of this issue, Bruchez *et al.* (*1*) developed a transposon screening approach in a human osteosarcoma cell line to identify a mechanism by which CD74, previously only associated with antigen presentation, directly inhibits EBOV and SARS-CoV-2 entry into host cells.

Viruses must gain entry into the host cell to replicate. In the case of EBOV, an enveloped virus, virions are internalized by macropinocytosis. Once virions reach endosomes, host cathepsin proteases cleave viral glycoproteins. The glycoproteins then fuse with the lysosomal membrane, which is followed by release of the viral genome into the host cell cytoplasm, where viral replication can occur (*2*). Thus, cathepsin-mediated cleavage is a critical step in the entry of many enveloped viruses, including EBOV, into the host cell.

Similar to EBOV, coronaviruses, including SARS-CoV-2, are enveloped viruses that undergo a series of entry steps culminating in genome release. Coronavirus entry also requires delivery of incoming viral particles to host lysosomes, where the coronavirus spike

Ebola virus cell entry

Normal cellular entry (left) of Ebola virus (EBOV) involves binding to cells expressing DC-SIGN (dendritic cell–specific ICAM-3–grabbing non-integrin 1) and TIM1 (T cell immunoglobulin mucin receptor 1), macropinocytosis, and cathepsin-mediated cleavage of the viral glycoproteins. Together with NPC1 (Niemann-Pick C1), glycoprotein cleavage allows fusion with endosomal membranes and genome release into the cytoplasm. However, CIITA (class II major histocompatibility complex transactivator) up-regulates the CD74 p41 isoform, which inhibits cathepsins and prevents genome release into the cytoplasm (right).



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Published by AAAS

protein is cleaved by cathepsins to facilitate fusion between virus and host membranes (*3*, *4*). However, in contrast to EBOV, SARS-CoV-2 also requires the activity of transmembrane serine protease 2 (TMPRSS2) to prime the viral spike protein (*5*). Thus, despite their differences in size and shape, EBOV and SARS-CoV-2 rely on similar proteolytic processes to gain entry into a target cell.

Bruchez *et al*. used a transposon screen in which transposable elements were inserted in front of or within genes. This approach allowed for tandem gene activation and inactivation in a single screen. To identify host factors involved in EBOV infection, the authors infected these cells with EBOV and identified two main "hits," including Niemann-Pick C1 (NPC1), an intracellular EBOV receptor that is required for entry, thus validating the approach (*6*). NPC1 is a cholesterol transporter in the lysosome and is essential for EBOV fusion of the glycoproteins with the lysosomal membrane and subsequent genome release.

Additionally, the authors found that activation of the transcription factor major histocompatibility complex (MHC) class II transactivator (CIITA) inhibited EBOV infection. CIITA is a nucleotide-binding oligomerization domain–like receptor (NLR). Typically, NLRs detect pathogen-associated molecular patterns (PAMPs) within the cell and trigger an intracellular antimicrobial signaling cascade leading to nuclear factor-κB (NF-κB) nuclear translocation and expression of various proinflammatory cytokines. Unlike most NLRs, CIITA acts mainly as a transcription factor to promote the expression of other genes, including serving as the master regulator of MHC gene expression. MHC presents peptides from either intracellular (MHC class I) or extracellular (MHC class II) proteins to adaptive immune cells. CIITA induces the expression of MHC class II genes to initiate antigen presentation. The authors determined that expression of CIITA was specifically associated with inhibition of cell entry by EBOV, thus defining the step of the viral life cycle that CIITA inhibits (see the figure).

Bruchez *et al*. identified CD74 as the CIITA-controlled host factor responsible for inhibiting EBOV entry. CD74, often called the invariant chain or Ii, is enriched in immune cell populations and associates with MHC class II. It localizes to endoplasmic reticulum (ER) membranes and facilitates MHC class II export from the ER to vesicles that fuse with the late endosome, resulting in trafficking to the cell surface (*7*). CD74 also blocks the peptide-binding groove so that MHC molecules do not bind peptides prior to trafficking. Thus, without CD74, MHC class II molecules are not properly processed, and antigen presentation becomes impaired.

Bruchez *et al*. show that the thyroglobulin domain of CD74 is required for its antiviral activity. This domain inhibits cathepsins (*8*). CD74 has four isoforms but only two of them, p41 and p43, have the thyroglobulin domain. The authors show that the p41 isoform is responsible for the antiviral activity of CD74 against EBOV entry and inhibits SARS-CoV-2 fusion, suggesting a broad antiviral activity of CD74 against many cathepsin-dependent viruses. These findings highlight the often shared strategies of distinct viruses that are co-opted from host cells to promote cell entry.

These findings suggest that molecules involved in antigen presentation could also possess direct antiviral activity and that other factors with defined functions may possess additional roles in antiviral immunity. CIITA activates antiviral factors that inhibit a broad range of viruses, such as human T cell leukemia virus type 2 (HTLV-2) (*9*), although the steps of the viral life cycle that it targets differ from those for EBOV and SARS-CoV-2. During HTLV-2 infection, CIITA acts more directly and inhibits the viral transactivator protein (TAX2), which promotes transcription of the viral genome and thus directly inhibits HTLV-2 replication (*9*). Some viruses have evolved mechanisms to inhibit this restriction. For example, Epstein-Barr virus (EBV), an oncogenic DNA virus, encodes Zta, a protein that directly inhibits CIITA and results in down-regulation of MHC class II molecules. This potentially allows EBV to escape recognition from the immune system (*10*).

The identification of host factors that could be targeted therapeutically to limit the replication of broad families of viruses may be an effective approach to combat viral-mediated disease. However, the therapeutic benefits of viral entry inhibitors are likely most effective prior to the onset of symptoms and the development of disease, given that by these stages, viral particles have already gained entry into the cell and begun to efficiently replicate.

REFERENCES AND NOTES

1. A. Bruchez *et al*., *Science* **370**, 241 (2020).
2. C. L. Hunt *et al*., *Viruses* **4**, 258 (2012).
3. J. Shang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11727 (2020).
4. T. Heald-Sargent, T. Gallagher, *Viruses* **4**, 557 (2012).
5. M. Hoffmann *et al*., *Cell* **181**, 271 (2020).
6. J. E. Carette *et al*., *Nature* **477**, 340 (2011).
7. O. Bakke, B. Dobberstein, *Cell* **63**, 707 (1990).
8. M. Mihelič *et al*., *J. Biol. Chem.* **283**, 14453 (2008).
9. C. Casoli *et al*., *Blood* **103**, 995 (2004).
10. D. Li *et al*., *J. Immunol.* **182**, 1799 (2009).

ACKNOWLEDGMENTS

A.I.W. is supported by NIH T32-AI060525 and NIH F31-AI149866, and C.B.C. is supported by NIH AI081759, AI150151, AI145828, and AI145296.

10.1126/science.abe2977

[1]Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [2]Center for Microbial Pathogenesis, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, USA. Email: coynec2@pitt.edu

NEUROSCIENCE

A new villain in neuronal death

A newly discovered partnership for glutamate triggers brain toxicity

By **Susan Jones**

Excitotoxicity is the process by which the excitatory amino acid neurotransmitter glutamate causes neuronal toxicity (*1*). A landmark study in 1987 revealed that Ca^{2+} influx into neurons through glutamate *N*-methyl-D-aspartate receptor channels (NMDARs) could trigger excitotoxicity (*2*), making Ca^{2+} ions potential villains as well as functionally critical signaling molecules in neurons. On page 191 of this issue, Yan *et al*. (*3*) reveal a newly discovered mechanism of NMDAR-mediated excitotoxicity: a physical interaction with TRPM4, a transient receptor potential channel (*4*). Small-molecule "interface inhibitors" prevent NMDAR-TRPM4 physical coupling and eliminate excitotoxicity in vitro and in vivo. This mechanism of excitotoxicity does not require NMDAR-mediated Ca^{2+} influx: TRPM4 is the new villain. Consequently, excitotoxicity can be halted without disrupting physiologically crucial NMDAR-mediated Ca^{2+} signaling. This presents opportunities in the design of therapeutics for disorders involving excitotoxicity, including stroke, epilepsy, and neurodegeneration.

Glutamate, released at excitatory synapses in the vertebrate brain, binds to NMDARs to induce Ca^{2+}-dependent forms of synaptic plasticity (*5*) and survival-promoting intracellular signaling pathways involving CREB [cyclic adenosine monophosphate (cAMP) response element–binding protein] (*6*). Glutamate can also spill over beyond the synapse and activate neighboring synaptic and extrasynaptic NMDARs. Excessive NMDAR-mediated Ca^{2+} influx is a widely accepted explanation for excitotoxicity (*7*, *8*).

A hitherto unrelated ion channel, TRPM4, is activated by cytoplasmic Ca^{2+} and is permeable to monovalent cations Na^+ and K^+ (but not to Ca^{2+}), mediating membrane depolarization (*9*). TRPM4 has previously been linked to neuronal death: In mice, inhibition of TRPM4 reduces neurodegeneration, and

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. Email: sj251@cam.ac.uk

Published by AAAS

genetic deletion of *Trpm4* protects against glutamate excitotoxicity (*10*).

Yan *et al.* demonstrate a physical interaction between NMDARs and TRPM4 in hippocampal cell cultures and brain tissue from mice. NMDARs are typically composed of GluN1, GluN2, and sometimes GluN3 subunits (*11*). TRPM4 specifically interacts with GluN2A and GluN2B, but not GluN2C/D, GluN1, or GluN3. The site of NMDAR-TRPM4 interaction was narrowed down to a cytoplasmic region of TRPM4, which the authors call TwinF. They further identify a highly conserved isoleucine-rich cytoplasmic sequence in GluN2A and GluN2B to which TwinF binds; they name this I4.

A TwinF peptide used to block NMDAR-TRPM4 interaction was neuroprotective in three assays of cell death: (i) NMDAR-evoked excitotoxicity of neuronal cultures, (ii) oxygen-glucose deprivation–evoked cell death in neuronal cultures, and (iii) a small but significant reduction in ischemic brain damage in vivo after middle cerebral artery occlusion (MCAO) in mice. Small-molecule inhibitors that target the precise NMDAR-TRPM4 interaction, called compounds 8 and 19, were identified. In the mouse MCAO model and in retinal degeneration induced by NMDA (the selective NMDAR agonist for which the receptor was named), intraperitoneal delivery of compound 8 offered neuroprotection—a small but significant reduction in infarct volume and cell loss, as well as reduction in NMDAR-TRPM4 interaction by as much as 38%. Thus, these small and simple molecules could potentially be delivered systemically to patients. Rapid delivery would be required in acute excitotoxic injury—for example, after stroke or epilepsy—whereas in chronic neurodegenerative diseases, possible contributions of excitotoxicity to disease progression (*8*) might be mitigated over a longer time course.

The authors hypothesize that NMDAR-TRPM4 coupling triggers "CREB shutoff" (CREB dephosphorylation) (*6*, *12*). TwinF and compounds 8 and 19 appear to "detoxify" NMDAR signaling by preventing CREB shutoff. In addition, TwinF reduces NMDAR-induced mitochondrial membrane dysfunction. By disrupting NMDAR-TRPM4 coupling, harmful signaling pathways are silenced while neuroprotective signaling pathways are spared (see the figure).

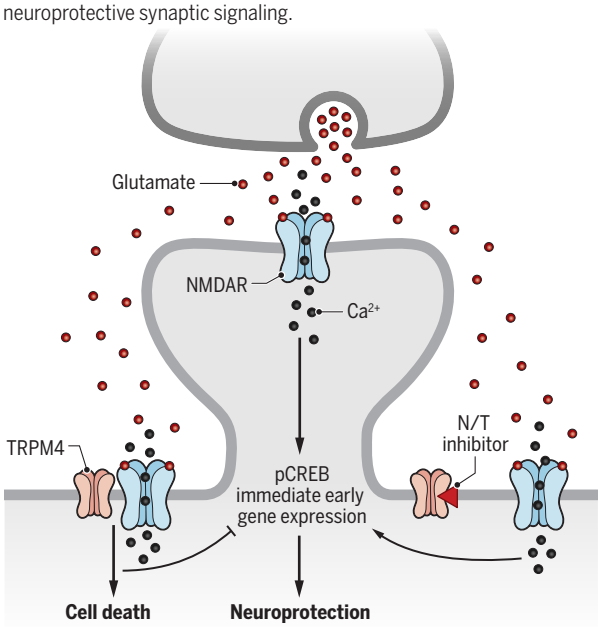
The mechanism by which physical coupling of NMDAR-TRPM4 brings about such neuroprotective effects is not yet elucidated, but it does not appear to involve Ca^{2+} signaling. A surprising finding is that blocking the NMDAR-TRPM4 interaction with TwinF or small-molecule inhibitors has no effect on normal synaptic NMDAR channel function recorded in hippocampal CA1 pyramidal neurons in mouse brain slices. Nor is there any effect on NMDAR-mediated Ca^{2+} influx and signaling in vitro. TRPM4 channel activity is also unaffected by blocking NMDAR interaction. Conventional NMDAR antagonists typically impair cognitive function (*13*), most likely by blocking NMDAR-mediated synaptic transmission, Ca^{2+} influx, and plasticity (*5*). Being able to target excitotoxicity while having no detrimental effect on NMDAR ionic current flux or calcium permeability in vitro constitutes an important step. It remains to be confirmed that the small-molecule inhibitors have no detrimental effects in vivo.

A physical coupling between TRPM4 and NMDARs in mediating excitotoxicity explains a conundrum in NMDAR research: why NMDARs located at synapses appear to mediate prosurvival signaling, whereas NMDARs located away from the synapse appear to prevent this and indeed trigger prodeath signaling pathways (*6*, *8*, *12*, *14*). Previous explanations for this include the subcellular localization of different signaling molecules or different concentrations of Ca^{2+} influx at synaptic versus extrasynaptic sites. Because TRPM4 is absent from synapses, this suggests that NMDAR-TRPM4 complexes would only form extrasynaptically, offering a satisfying explanation for this puzzle.

Considerable research effort has focused on the role of Ca^{2+} in NMDAR-mediated excitotoxicity. The study of Yan *et al.* introduces a new villain, TRPM4. This requires reevaluation of the role of NMDARs in neuronal death. It may redirect research efforts away from Ca^{2+} and toward physical interactions of NMDARs, and it could provide another argument against developing NMDAR antagonist and channel blocker therapies. It is unknown whether the NMDAR-TRPM4 interaction will prove to be the primary (or even exclusive) excitotoxicity mechanism in a broader range of neurons. For example, neurodegeneration in Huntington's and Parkinson's diseases affects the basal ganglia (*14*), whereas Yan *et al.* studied hippocampal neurons. There is evidence for acute and chronic excitotoxicity in human neurological disorders (*8*), and so it may be illuminating to screen postmortem human brains from patients with neurodegenerative diseases for NMDAR-TRPM4 interactions. If higher amounts of interaction are seen compared with healthy subjects, can small-molecule protein interaction inhibitors be designed to produce a sufficiently robust yet safe disruption? Another intriguing question is whether NMDAR-TRPM4 interaction is triggered under normal physiological conditions and plays an important functional role, or is instead purely a pathological response to injury. After more than 30 years of research, perhaps Ca^{2+} signaling will ultimately be shown to have little importance in excitotoxicity—to be a hero and not a villain in neuronal signaling. ■

Disrupting glutamate receptor interactions

Synaptic *N*-methyl-D-aspartate receptor channels (NMDARs), activated by glutamate release and coincident membrane depolarization, trigger neuroprotective signaling pathways involving phosphorylated cAMP-responsive element–binding protein (pCREB). Extrasynaptic NMDARs that are activated by synaptic glutamate spillover interact with transient receptor potential monovalent cation channel 4 (TRPM4), triggering excitotoxic cell death. Blocking NMDAR-TRPM4 binding with an N/T inhibitor eliminates neuronal death signaling and enables neuroprotective synaptic signaling.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. J. W. Olney, *Science* **164**, 719 (1969).
2. D. W. Choi, *J. Neurosci.* **7**, 369 (1987).
3. J. Yan *et al.*, *Science* **370**, eaay3302 (2020).
4. P. Launay *et al.*, *Cell* **109**, 397 (2002).
5. C. Lüscher, R. C. Malenka, *Cold Spring Harb. Perspect. Biol.* **4**, a005710 (2012).
6. G. E. Hardingham, H. Bading, *Nat. Rev. Neurosci.* **11**, 682 (2010).
7. M. Arundine, M. Tymianski, *Cell Calcium* **34**, 325 (2003).
8. J. Lewerenz, P. Maher, *Front. Neurosci.* **9**, 469 (2015).
9. R. Guinamard, C. Simard, L. Sallé, in *Encyclopedia of Signaling Molecules*, S. Choi, Ed. (Springer, 2016).
10. B. Schattling *et al.*, *Nat. Med.* **18**, 1805 (2012).
11. D. J. A. Wyllie, M. R. Livesey, G. E. Hardingham, *Neuropharmacology* **74**, 4 (2013).
12. G. E. Hardingham, Y. Fukunaga, H. Bading, *Nat. Neurosci.* **5**, 405 (2002).
13. S. A. Lipton, *Nat. Rev. Drug Discov.* **5**, 160 (2006).
14. M. P. Parsons, L. A. Raymond, *Neuron* **82**, 279 (2014).

10.1126/science.abe2791

Published by AAAS

QUANTUM MATERIALS

Layer-cake 2D superconductivity

An easy-to-synthesize quantum material may help unveil quantum effects

By **Leslie M. Schoop**

Materials science historically has had an immense impact on humanity, as entire periods of time, such as the Bronze or Iron ages, have been named after materials (*1*). In the modern era, a new class of materials called quantum materials has started to affect people's lives. Quantum materials are broadly defined as materials whose electronic or magnetic behavior cannot be explained by classical physics (*2*). Discoveries of quantum materials with distinct or improved properties are often followed by a surge of research, resulting in either the discovery of new physics, or in applications, such as in low–power consumption electronics, sensing, photodetection, high-speed electronics, or quantum information science (*3*). On page 231 in this issue, Devarakonda *et al.* report the synthesis of a highly interesting quantum material (*4*), which may facilitate the study of quantum effects that so far have been obscured.

Superconductors are a class of fascinating quantum materials with properties that inspired dreams of applications without power loss, such as high-speed, levitating trains. These dreams have in part been realized through development of materials that superconduct at high enough temperatures for cooling of liquid nitrogen instead of helium, which is expensive and sparse (*5*). The discovery of these high–critical temperature superconductors created a fascinating puzzle. The superconductivity of these materials is "unconventional," meaning that the mechanism is not fully understood. One way to shine light on the physics behind the superconductivity is to reduce the dimensionality of the materials. Two-dimensional (2D) superconductors simplify the problem and also offer access to many other unsolved mysteries of condensed matter physics. These include the Berezenskii-Kosterlitz-Thouless (BKT) transition (*6*), which is an early example of a "topological" phase transition whose discovery was awarded the 2016 Nobel Prize in Physics.

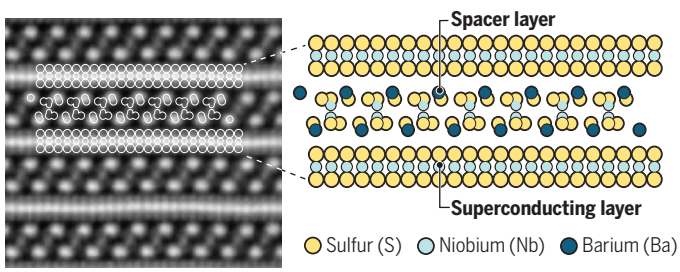
Although materials that exhibit 2D superconductivity are known, imperfections in the crystals have impeded study. Because 2D materials are only a few atoms thick, they often suffer from degradation upon exposure to air or from defects that result from exfoliation. For these reasons, 2D materials are usually manually encapsulated with protecting layers, typically boron nitride (*7*). To create a system for studying the pure physics of 2D superconductivity, a known material can be made as pure as possible, or a new material can be designed. Although purifying materials is costly and time-consuming, developing new materials presents a host of other challenges. The discovery of clean 2D superconductivity in $Ba_6Nb_{11}S_{28}$ by Devarakonda *et al.* opens the door to a better understanding of 2D superconductivity and its associated quantum phenomena.

The material is an intrinsic heterostructure consisting of alternating layers of the 2D superconductor NbS_2 and an electronically uninteresting spacer layer, Ba_3NbS_5 (see the figure). Think of the material as a layer cake with a thin layer of chocolate repeatedly inserted between thicker layers of cake. The chocolate is the superconductor (NbS_2) and the dough is the spacer layer that protects the chocolate from cracking or being attacked by air or moisture. Because of this protection, the NbS_2 layer exhibits much cleaner 2D superconductivity than a single, unprotected layer. This feature is manifested by the mobility of electrons within the protected NbS_2 layer, which is three orders of magnitude higher than what has been reported in the unprotected counterpart. In addition, the BKT transition can be observed in $Ba_6Nb_{11}S_{28}$ because of the high purity of the NbS_2 layers.

The beauty of this material also lies in the natural growth of the heterostructure, as if the aforementioned layer cake could be baked by merely mixing the chocolate and batter in a bowl and putting the mixture in an oven with the two components naturally separating during the baking process. This makes the synthesis process much less labor-intensive than adding each layer manually, as is necessary when protecting 2D crystals with boron nitride. Because of the ease of synthesis, different types of layered materials may be developed where the 2D layers are naturally protected by their environment. The 2D layer would not always have to be a superconductor. Different types of quantum materials are also possible. One example might be topological insulators, which belong to a class of quantum materials with promise for several applications, one example being quantum computing (*8*). The authors' discovery suggests a much simpler alternative to exfoliated nanodevice fabrication. The key question is whether this strategy will easily extend to other materials beyond $Ba_6Nb_{11}S_{28}$. ■

Protected 2D layers

Alternating layers of superconducting NbS_2 and a Ba_3NbS_5 spacer allow high electron mobility in the NbS_2 while also protecting it. This creates a "layer cake"– like structure that permits clean superconducting behavior.



GRAPHIC: N. DESAI/*SCIENCE* ADAPTED FROM (*4*)

Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. Email: lschoop@princeton.edu

REFERENCES AND NOTES

1. N. A. Spaldin, *VSH-Bull.* no. 2 (August 2017), p. 11.
2. B. Keimer, J. E. Moore, *Nat. Phys.* **13**, 1045 (2017).
3. Y. Tokura, M. Kawasaki, N. Nagaosa, *Nat. Phys.* **13**, 1056 (2017).
4. A. Devarakonda *et al.*, *Science* **370**, 231 (2020).
5. R. J. Cava, *J. Am. Ceram. Soc.* **83**, 5 (2000).
6. A. M. Goldman, in *40 Years of Berezinskii–Kosterlitz–Thouless Theory* (World Scientific, 2013), pp. 135–160.
7. K. S. Novoselov *et al.*, *Science* **353**, aac9439 (2016).
8. S. D. Sarma, M. Freedman, C. Nayak, *npj Quantum Inf.* **1**, 1 (2015).

ACKNOWLEDGMENTS

The author acknowledges support by the Gordon and Betty Moore Foundation through grant GBMF9064.

10.1126/science.abd4225

Published by AAAS

ELECTRON MICROSCOPY

Better, faster, and even cheap

An inexpensive gold grid with submicrometer holes minimizes protein sample movement

By **Micah Rapp** *and* **Bridget Carragher**

Cryo–electron microscopy (cryo-EM) enables access to structures of proteins that were previously intractable, including large protein complexes such as the ribosome (*1*), integral membrane proteins (*2*, *3*), and highly heterogeneous or conformationally dynamic systems (*4*). Each sample is a vitrified layer of protein suspended over a support film on an EM grid. Despite recent advances in cryo-EM (the so-called "resolution revolution") (*5*, *6*), major barriers persist, including loss of the highest-resolution information through electron beam damage and blurring from sample movement (which is most pronounced initially when the sample is least damaged). Typically, tens of thousands of images must be averaged to compensate for signal loss. On page 223 of this issue, Naydenova *et al.* (*7*) describe a new specimen support film (see the figure) that not only improves both the quality of images and the efficiency of collection, but also does so at a relatively low price.

Like much of society, the strengths and weaknesses of cryo-EM have been highlighted by the coronavirus disease 2019 (COVID-19) pandemic. Structural biology has contributed to our understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with more than 330 structures deposited into the Protein Data Bank since early March. At the time of writing, 264 of those structures were produced by x-ray crystallography and 72 by cryo-EM. The astonishing throughput of crystallography was demonstrated in recent work on the main SARS-CoV-2 protease, where ~1500 datasets were collected in a single day (*8*).

In contrast, cryo-EM datasets typically require on the order of 24 hours of data collection, even though the process is largely automated (*9*). High-end microscopes are not only under extraordinary demand, but are also associated with very high capital and annual maintenance costs (*10*). These differences have made x-ray crystallography better suited to the rapid turnaround required by the pharmaceutical drug discovery pipeline (*11*). These timelines also present a barrier for academic research. Increasing throughput will lower costs and enable more researchers to use cryo-EM to solve structures and accelerate scientific discovery.

Naydenova *et al.* propose a new specimen support film, dubbed "HexAuFoil," that provides many advantages over conventional support films. In cryo-EM, images are typically taken of proteins embedded in a layer of vitrified ice suspended over a support films containing holes ~1 to 2 μm in diameter. By making these holes smaller (200 to 300 nm in diameter) and packing them more tightly together (a nontrivial feat), they substantially increased data throughput. They obtained 200 images for every microscope stage movement (versus a typical 10 to 30 images) and propose that as many as 800 images could be obtained with an appropriately configured microscope.

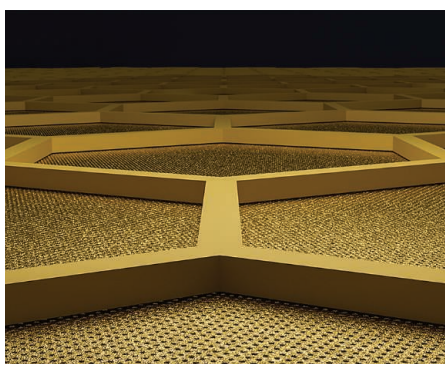
A second approach to increasing throughput that Naydenova *et al.* addressed is to increase the data quality by minimizing the information loss from both radiation damage and sample movement. They present a fascinating and exhaustive analysis showing that the primary cause of specimen movement during imaging is a buckling of the vitrified ice layer. On the basis of these insights, they created a support film with an optimal substrate thickness relative to the hole diameter, which reduces the total movement of the sample to <1 Å during the course of an exposure. They could mathematically extrapolate the data to a three-dimensional map before the onset of radiation damage.

Standard support films have recently been used to obtain the first truly atomic-level cryo-EM reconstructions (*12*, *13*) of apoferritin, a very stable test specimen. The advances developed by Naydenova *et al.* should bring this goal closer for less well-behaved proteins. Although software is also now available to correct for the effects of beam-induced movement (*14*), this approach does not provide the concomitant benefit of faster data collection.

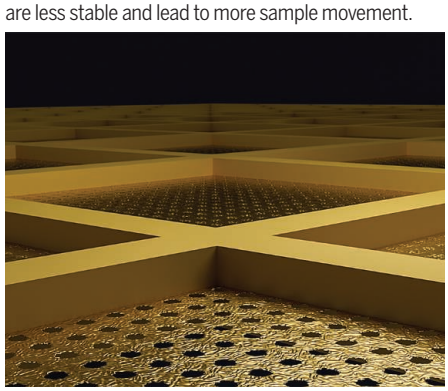
High-resolution cryo-EM is still in a state of active method development, its true potential not yet realized. Thus, it is exciting that Naydenova *et al.*, with one accessible, inexpensive hardware development, will allow all practitioners to acquire better images much more rapidly as soon as the grids become commercially available. The promise is of a future of high-resolution structures of a wide range of proteins, in an ensemble of conformational or compositional states (*15*), produced with much higher throughput. ■

Grid schematics

The sample holder developed by Naydenova *et al.* for cryo-EM has a dense packing of holes 200 to 300 nm in diameter for supporting vitrified proteins.



Conventional holders with holes 1 to 2 μm in diameter are less stable and lead to more sample movement.



ILLUSTRATIONS: MICAH RAPP

Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY 10027, USA, and Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. Email: bcarr@nysbc.org

REFERENCES AND NOTES

1. A. Brown, S. Shao, *Curr. Opin. Struct. Biol.* **52**, 1 (2018).
2. Y. Cheng, *Curr. Opin. Struct. Biol.* **52**, 58 (2018).
3. M. Dong *et al.*, *Nat. Commun.* **11**, 4137 (2020).
4. T. Nakane, D. Kimanius, E. Lindahl, S. H. W. Scheres, *eLife* **7**, e36861 (2018).
5. W. Kühlbrandt, *Science* **343**, 1443 (2014).
6. Y. Cheng, *Science* **361**, 876 (2018).
7. K. Naydenova, P. Jia, C. J. Russo, *Science* **370**, 223 (2020).
8. A. Douangamath *et al.*, bioRxiv 118117 (2020).
9. D. Lyumkis, *J. Biol. Chem.* **294**, 5181 (2019).
10. K. Naydenova *et al.*, *IUCrJ* **6**, 1086 (2019).
11. G. Scapin, C. S. Potter, B. Carragher, *Cell Chem. Biol.* **25**, 1318 (2018).
12. T. Nakane *et al.*, bioRxiv 110189 (2020).
13. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, bioRxiv 106740 (2020).
14. D. Tegunov, L. Xue, C. Dienemann, P. Cramer, J. Mahamid, bioRxiv 136341 (2020).
15. A. Dance, *Nat. Methods* **17**, 879 (2020).

ACKNOWLEDGMENTS

Supported by NIH grants GM103310 and OD019994 and Simons Foundation grant SF349247.

10.1126/science.abd8035

Published by AAAS

VIEWPOINT: COVID-19

Contextualizing bats as viral reservoirs

Preventing zoonotic emergence from bats requires integrative research

By **Daniel G. Streicker**[1,2] *and* **Amy T. Gilbert**[3]

Coronavirus disease 2019 (COVID-19) is the latest in a distressing tally of viral infections—including Ebola, Nipah, rabies, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS)—that have evolutionary origins or epidemiological associations with bats. This seeming preponderance of zoonoses has propelled bats from biomedical obscurity to the forefront of global health. Immunological traits have been proposed to allow bats to control viruses differently from other animals. However, incomplete baselines for broader comparisons across vertebrates and extensive immunological variation among bat species casts uncertainty on their distinctiveness as viral reservoirs. Moreover, common perceptions that bats asymptomatically harbor viruses more often than other animals and that their viruses are more diverse or pose systematically heightened zoonotic risk remain unresolved. The search for answers may inspire new approaches to manage disease threats to human and animal health.

Bats (order Chiroptera) comprise ~1400 species that split from the remaining members of the Scrotifera (carnivores, pangolins, cetaceans, and odd- and even-toed ungulates) over 60 million years ago. The capacity for true flight, specific to bats among mammals, opened diverse trophic niches, making bats key providers of global ecosystem services, including insect pest control, seed dispersal, and pollination of agricultural plants. Flight also introduced physiological challenges that transformed bat life history. For example, aerial transport of young restricts litter sizes to one or two pups annually across most species. The need for multiple bouts of reproduction to maximize fitness therefore favored longevity, hypothesized to be mediated by adaptations to suppress tumors and inflammation caused by DNA damage (*1*).

Perhaps serendipitously, these mechanisms also limit virus-induced inflammation, potentially explaining why viruses including Marburg virus, SARS-coronavirus (SARS-CoV), and MERS-CoV are thought to cause subclinical infections in the presumed natural bat hosts (Egyptian fruit bats, *Rousettus aegyptiacus*, for Marburg virus and horseshoe bats, *Rhinolophus* spp., for both CoVs) but immunopathology in other vertebrates. Over evolutionary time scales, limited inflammatory responses in bats, together with high population densities and gregarious social behaviors in some species that may facilitate virus transmission, could have selected for viruses that cause severe disease in incidental hosts that lack analogous defenses.

Peculiarities in bat immune systems that plausibly alter viral interactions are increasingly recognized (*2*). Whether bats are exceptional in this respect is unclear because knowledge of vertebrate immune systems largely derives from inbred mice or immortalized cells, which diverge substantially from wild relatives. Fortunately, the rise in genome sequencing has provided crucial phylogenetic context to the evolutionary origins of bat immunity while facilitating comparisons with diverse nonmodel species (*3*). For example, comparative transcriptomics showed distinct aspects of innate immunity in little brown bat (*Myotis lucifugus*) and large flying fox (*Pteropus vampyrus*) but also in eight other mammalian and avian species (*4*). By characterizing distinct antiviral features across taxa, efforts to contextualize bat immunity might inspire new strategies to prevent and treat viral zoonoses in humans and animals.

Heightened interest in bat-associated viral zoonoses has also revealed high immunological variation among species. For example, black flying foxes (*P. alecto*) have an unusually contracted interferon-α (IFN-α) locus (genes that encode components of the innate immune response) and cells that constitutively express IFN-α, inducing antiviral activity (*5*). However, other bat species have expanded IFN-α loci and lack constitutive IFN-α (*6*). Similarly, bat species with increased constitutive antiviral defenses may do so through differing gene expression pathways (*4*), and the antiviral APOBEC gene family has undergone bat lineage–specific expansions or duplication (*3*). This implies that some of the unusual antiviral defenses in bats arose independently after the evolution of flight. Divergent immunological repertoires among bat species may reflect alternative responses to biogeographic variation in viral assemblages and environmental conditions. Identifying the eco-evolutionary determinants and range of antiviral defenses might help identify unreported reservoirs of zoonoses but requires expanding research beyond the relatively few bat species known to transmit zoonoses.

Whether features of bat immunology predictably translate into functionally distinct antiviral strategies is unresolved. For example, the popular notion that bats tolerate virus infections is supported by experimental infections of bats with Marburg virus, Ebola virus, and MERS-CoV. Conversely, other viruses that may be lethal to humans—including lyssaviruses, Tacaribe virus, and Lloviu virus (human pathogenicity unknown)—are also apparently lethal to bats, including putative reservoir hosts. Sublethal effects of viruses on wild bats are largely undetectable because longitudinal monitoring of individuals is only possible in philopatric species, which live in relatively small groups and can be reliably recaptured. Individual heterogeneities that alter infection outcomes in humans and other animals—such as age, sex, social hierarchies, and past and contemporaneous infections—remain virtually unexplored in bats. Given limited evidence from wild populations, meta-analyses of experimental infections might test whether bats systematically manifest less symptomatic disease than other hosts. Other taxa that are infected with some zoonotic viruses but exhibit mild or no symptoms, such as rodents (for example, Lassa virus) and birds (for example, West Nile virus), provide relevant contrasts.

Whether bat viruses are disproportionately zoonotic is an outstanding global health conundrum. A meta-analysis of 2805 host-virus interactions showed that bats are more likely than other mammals to be infected by viruses that also infect humans (*7*). Yet when analyses are restricted to hosts that are believed to be important for natural transmission cycles, viral richness among Chiroptera was unexceptional, and they contributed approximately the number of zoonoses expected for the number of species in this order (*8*). Thus, evolutionarily conserved traits of bats seem unlikely to produce viruses with inflated zoonotic capability. Heightened susceptibility or perhaps surveillance may explain why bats appear to host a relatively large number of zoonotic viruses.

Once introduced into the human population, are bat viruses exceptionally dangerous?

[1]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [2]Medical Research Council–University of Glasgow Centre for Virus Research, Glasgow, UK. [3]U.S. Department of Agriculture, Animal and Plant Health Inspection Service, National Wildlife Research Center, Fort Collins, CO, USA. Email: daniel.streicker@glasgow.ac.uk; amy.t.gilbert@usda.gov

Published by AAAS

One meta-analysis found higher case fatality ratios (CFRs) and lower human-to-human transmissibility of bat viruses; however, the extent that these patterns generalize among bat viruses was uncertain (*9*). The rabies-causing lyssaviruses, which comprise ~50% of zoonotic viruses recognized from bats (*8*), exemplify high CFRs and low transmissibility among humans but, being lethal across all mammals, do not fit the emerging paradigm of tolerance in bats contrasted with virulence in humans. Deviations such as SARS-CoV-2 (low CFR and high transmissibility) and the ebolaviruses (moderate CFR and transmissibility) highlight further complexity.

If the virulence of bat viruses is systematically increased, mathematical models fit to in vitro experiments provide a possible explanation: Accelerated viral propagation with limited cellular morbidity might favor chronic subclinical infections in bats but acute infections in other hosts (*10*). Although the prediction that bat viruses that cause short-lived, lethal infections in humans infect bats chronically remains unconfirmed in vivo, the short time frames and small sample sizes of most experiments make detecting reactivation of latent viral infections in bats unlikely. Ultimately, virulence is an emergent property of host and virus interactions. As such, determining whether differences among species arise from virus-specific phenomena within bats, inappropriate responses of naïve immune systems, or viral tolerance mechanisms may require profiling immunological responses and within-host dynamics across diverse viruses and host species.

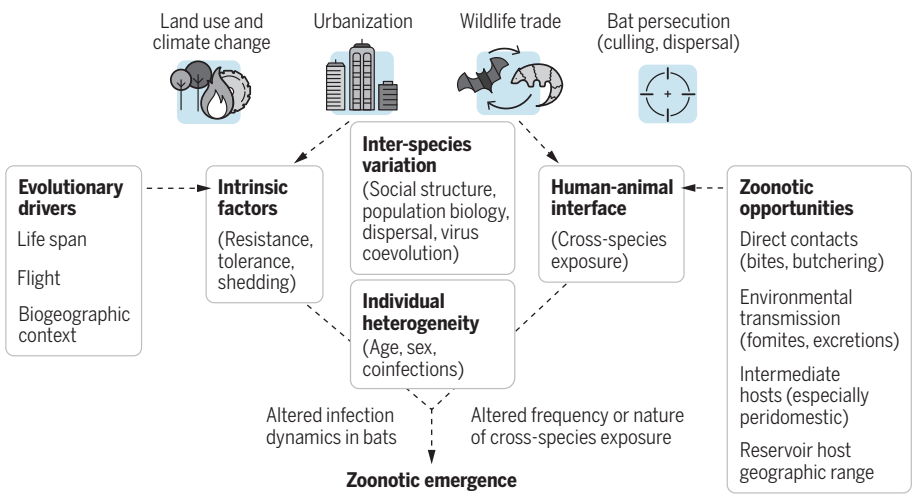
Beyond contextualizing the distinctiveness of bat reservoirs, research should also tackle the real-world complexity underlying viral zoonotic emergence (see the figure). A first step may be to identify how intrinsic traits of bats and extrinsic factors interact to govern viral transmission, community assembly, and zoonotic emergence. For example, spatially replicated metagenomic sequencing in vampire bats (*Desmodus rotundus*) found no evidence that larger colonies sustain more viruses but revealed elevational gradients and age biases in viral diversity (*11*). In flying foxes, longitudinal monitoring showed pulsed shedding of multiple paramyxoviruses (a virus family associated with several emerging zoonoses), potentially arising from physiological stress induced by acute food shortages (*12*). Understanding virus coinfection and community dynamics may also reveal recombination opportunities that potentially enable emergence. Anticipating how anthropogenic perturbations such as land use change, persecution of bats for consumption, trade, or from fear, or misguided efforts at disease control will precipitate emergence is a greater challenge. These actions can alter both viral transmission among bats and the frequency of interspecies contacts (including with intermediate hosts) but are conceptually underdeveloped and rarely tested empirically.

Integrating understanding of the zoonotic process across biomedical, population, and ecosystem scales may enable prevention of zoonotic emergence by reducing virus circulation in bat reservoirs. Knowledge of bat genomics and immunity opens the door for using genetic editing technologies such as CRISPR to engineer viral resistance in wild bats, analogous to ongoing efforts to control Lyme disease in wild mice (*13*). Historical barriers to delivering vaccines at sufficient scales to alter viral dynamics in wild bat populations are also diminishing. The relative ease of metagenomic sequencing enables rapid discovery of naturally occurring, innocuous, and species-specific bat viruses that might be engineered into transmissible vaccines targeting zoonoses in wild bats. This idea has empirical precedents from efforts to vaccinate wild rabbits against myxomatosis and rabbit hemorrhagic disease (*14*). The traits expected to facilitate virus transmission in some bats, such as gregariousness and flight, could support live vaccine dissemination, and naturally slow demographic turnover would help to maintain vaccine-induced population-level immunity, allowing less frequent interventions (*15*). Such potentially transformative strategies require rigorous investigation into efficacy, safety, and ecological impacts as well as overcoming barriers to societal acceptance. Viruses such as rabies virus, Marburg virus, and henipaviruses—for which bat reservoirs are known, host and viral genomes are available, and transmission to humans and/or animals occurs with measurable frequency—can serve as tractable and important models to evaluate and refine candidate interventions.

Viral emergence from bats is largely unpredictable and unpreventable. Solutions require qualitative and quantitative expansions over current practice in bat research, which rarely considers heterogeneities among individuals, populations, and species. This variability can reveal the drivers and phenotypic importance of bat-virus interactions as well as whether they generalize in ways that might aid surveillance or management of zoonotic threats. Given the costs of the COVID-19 pandemic, the need for an ambitious research agenda is more evident now than ever. ■

Dimensions of zoonotic emergence

Bat viruses emerge through currently unpredictable interactions of evolutionary and ecological forces. Intrinsic features of bat immune systems have been shaped by bat life history and past viral interactions. Anthropogenic perturbations may alter host-virus interactions at the individual or population levels while breaking down historical barriers between species, culminating in viral emergence.



GRAPHIC: JOSHUA BIRD/*SCIENCE*

REFERENCES AND NOTES

1. G. Zhang *et al.*, *Science* **339**, 456 (2013).
2. J. N. Mandl *et al.*, *Front. Immunol.* **9**, 2112 (2018).
3. D. Jebb *et al.*, *Nature* **583**, 578 (2020).
4. A. E. Shaw *et al.*, *PLOS Biol.* **15**, e2004086 (2017).
5. P. Zhou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2696 (2016).
6. S. S. Pavlovich *et al.*, *Cell* **173**, 1098 (2018).
7. K. J. Olival *et al.*, *Nature* **546**, 646 (2017).
8. N. Mollentze, D. G. Streicker, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9423 (2020).
9. S. Guth *et al. Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20190296 (2019).
10. C. E. Brook *et al.*, *eLife* **9**, e48401 (2020).
11. L. M. Bergner *et al.*, *Mol. Ecol.* **29**, 26 (2020).
12. A. J. Peel *et al.*, *Emerg. Microbes Infect.* **8**, 1314 (2019).
13. J. Buchthal *et al. Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20180105 (2019).
14. J. J. Bull *et al.*, *Trends Microbiol.* **26**, 6 (2018).
15. K. M. Bakker *et al.*, *Nat. Ecol. Evol.* **3**, 1697 (2019).

ACKNOWLEDGMENTS

D.G.S. is funded by a Wellcome Trust Senior Research Fellowship (217221/Z/19/Z). We thank M. Palmarini, S. Babayan, and M. Viana for discussions.

10.1126/science.abd4559

Published by AAAS

POLICY FORUM

SPACE POLICY

U.S. policy puts the safe development of space at risk

Promoting national regulation of space mining rather than multilateral governance risks a "race to the bottom"

By **Aaron Boley**[1] *and* **Michael Byers**[2]

In September 2020, the U.S. National Aeronautics and Space Administration (NASA) announced that it is seeking proposals from private companies to extract small amounts of regolith from the surface of the Moon. According to NASA Administrator Jim Bridenstine, the exercise will buttress an interpretation of international space law that is favored by the U.S. government: "What we're trying to do is make sure that there is a norm of behavior that says that resources can be extracted and that we're doing it in a way that is in compliance with the Outer Space Treaty" (OST) (*1*). NASA's move is part of a larger U.S. diplomatic effort. In April 2020, President Trump signed an executive order (EO) that affirmed a right of commercial space mining and directed the U.S. State Department to secure the expressed support of U.S. allies. Shortly after, NASA announced a plan for bilateral Artemis Accords, which, if accepted by many nations, could enable the U.S. interpretation of international space law to prevail and make the United States—as the licensing nation for most of the world's space companies—the de facto gatekeeper to the Moon, asteroids, and other celestial bodies. Because acquiescence is often treated as consent in international law, even NASA's purchase of regolith would, if not protested by other nations, strengthen the U.S. interpretation. Other nations need to speak up, now.

The Artemis Accords are to include recognition of a right to commercial space mining subject to national regulation only (i.e., no need for a new multilateral agreement), as well as the right of companies to declare "safety zones" around their operations to exclude other actors. Negotiations with individual NASA partner nations are reportedly under way, with the conclusion of an Artemis Accord being a prerequisite for participation in NASA's Artemis lunar program. For many nations, the chance to have their own astronaut on the Moon is a powerful incentive.

The permissibility of commercial space mining has become a central aspect of U.S. space policy: In 2015, the Commercial Space Launch Competitiveness Act gave U.S. citizens and companies the right to own and sell space resources under U.S. law. In 2017, Luxembourg followed suit, adopting legislation in support of commercial space mining and offering subsidies to space mining companies that incorporate there. Other nations might seek to attract companies by offering minimal regulations and lax enforcement. Leaving regulation to national governments thus risks a "race to the bottom" and even the emergence of "flags of convenience."

The 2020 EO goes further than the 2015 legislation by explicitly rejecting space as a "global commons" and dismissing the 1979 United Nations (UN) Moon Agreement as irrelevant because it has not been ratified by major spacefaring nations. The Moon Agreement declares the Moon and other celestial bodies the "common heritage of mankind" and provides a mechanism for initiating a multilateral negotiation on space mining.

Most other nations take the view that space mining should only be conducted under an international system of regulation, oversight, and benefit sharing. The 1967 OST provides the governance framework for all space activities. Article II reads: "Outer space, including the moon and other celestial bodies, is not subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means."

The term "national appropriation," however, is not defined in the OST, making it possible to argue that the extraction and even the sale of space resources is not prohibited and is therefore already permitted. NASA's Bridenstine has drawn an analogy to fishing on the high seas, where a fish cannot be owned while in the ocean but can be owned as soon as it is caught. The analogy is apt because it concerns the acquisition of ownership of something from an "area beyond national jurisdiction." Still, it should be noted that fishing without science-based regulation often leads to overexploitation and even destruction of stocks, and that nations have responded to these risks by negotiating multilateral treaties on the management of high-seas fisheries that result in science-based quotas and sometimes moratoria.

The Artemis Accords have been condemned by Russia as a blatant move to remake international space law in favor of the United States. The first casualty of this initiative might therefore be the long and stable relationship of space cooperation between these two nations. Nor is there any indication that the United States will involve China, another superpower and an increasingly capable space actor, in this effort to develop a new governance regime.

MINING OPPORTUNITIES AND RISKS

The commercial potential of space mining is receiving most of the attention, but there is also a strong motivation rooted in science and exploration. At least 14 national space agencies have identified in situ resource use (ISRU) as a needed capability for long-duration missions, including crewed missions to the Moon, Mars, and deep space (*2*). Artemis will be the first such NASA-led program (*3*). Resources such as ice and water-bearing minerals from the lunar South Pole will provide fuel, radiation shielding, and life support for surface and orbital operations. The regolith will be mined for construction materials and as a source of hydrogen and oxygen (*4*). Many asteroids also contain an abundance of water and minerals (*5*, *6*) that could be used to support space operations.

ISRU will provide new science opportunities and unprecedented sampling of celestial bodies. For example, asteroids contain some of the oldest materials in the Solar System, some of which have experienced little thermal processing since their incorporation into parent bodies. The Moon's ice deposits are a partial record of volatile delivery to Earth. However, space mining, especially if conducted by loosely regulated private companies, could hinder science. For example, water and oxygen could be extracted from astromaterials by pyrolysis (*7*), and, without systematic scientific sampling before alteration or consumption, valuable information about the Solar System (e.g., locked into cosmochemical or mineralogical signatures) could be lost. Analysis to maximize resource yields is not generally the same as that needed for understanding the Solar System. Inconsistent practices in resource extraction, a likely result of purely national regulations, would only exacerbate the losses in scientific opportunities (*8*).

[1]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada. [2]Department of Political Science, University of British Columbia, Vancouver, BC, Canada. Email: acboley@phas.ubc.ca

Published by AAAS

Some of the first efforts at private space exploration have already manifested a less than rigorous approach to risk avoidance. In 2019, the Israeli nonprofit SpaceIL crashed a robotic lander on the Moon. Unbeknownst to SpaceIL, its partner—the Arch Mission Foundation—had placed thousands of nearly indestructible tardigrades on board (*9*). In 2018, SpaceX launched a Tesla automobile on an orbit that extends past Mars; although no impact with Mars is expected, there was an initial lack of clarity on the mission profile and the potential for the unsterilized payload to encounter Mars (*10*).

Mining can generate serious operational concerns. Lunar dust is a known challenge to operations on the Moon. Any surface activity could exacerbate lunar dust migration, including by lofting dust onto trajectories that cross lunar orbits, such as that of NASA's proposed Lunar Gateway (*11*). Moreover, without cooperation by all actors, the limited number of useful lunar orbits could quickly become filled with space debris.

On asteroids, low escape speeds will make it difficult to prevent the loss of surface material. Even if full enclosures are used, waste material may be purposefully jettisoned. Mining could also lead to uncontrolled outbursts of volatile sublimation after the removal of surface layers. Because the asteroids targeted for mining are likely to be those with small minimum orbit intersection distances, the resulting meteoroid debris streams could threaten lunar operations as well as satellites in Earth's orbit (*12*). In a worst-case scenario, a trajectory change resulting from mining could eventually lead to an Earth-impact emergency.

Space missions already provide some evidence of these risks. In 2019, during the course of Japan's Hayabusa2 mission, a small impactor was used to make a crater on (162173) Ryugu (*13*). Some of the resulting anthropogenic meteoroids could begin reaching Earth during the 2033 apparition. In 2022, NASA will test its ability to deflect an asteroid by striking (65803) Didymos B (Dimorphos) with the Double Asteroid Redirection Test spacecraft. This impact will produce anthropogenic meteoroids, with the possibility of immediate delivery to Earth (*14*). Although these risks are small, they demonstrate how easily human actions can change the near-Earth environment.

As for the safety zones envisaged in the Artemis Accords, these would provide a space actor with many of the benefits of territory while relieving it of long-term obligations of stewardship. There are analogies here on Earth, including "orphan" oil wells and abandoned mine tailings.

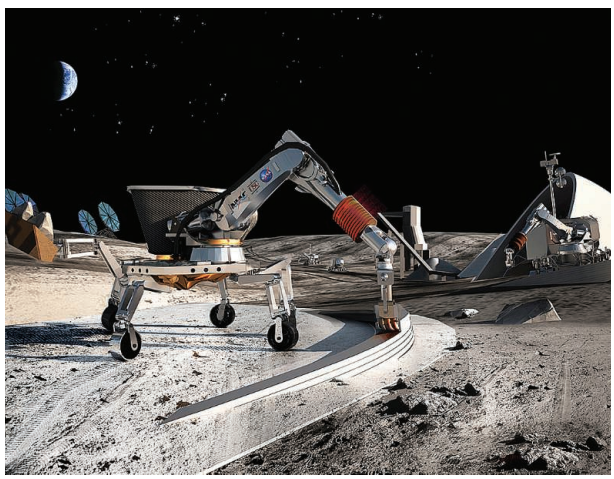
MANAGING A GLOBAL COMMONS

Lunar dust, anthropogenic meteoroids, and the loss of science opportunities are examples of negative externalities that could arise because of space mining. Their inadequate consideration can be expected when rulemaking and oversight in an area beyond national jurisdiction are left to individual nations rather than a multilateral regime. The approach being taken by the U.S. government, of bilateral negotiations and purely national regulations, ignores widely adopted insights into the management of common pool resources and global commons (*8*, *15*). Multilateral management takes time and requires compromise but also internalizes externalities by, for example, ensuring a degree of peer review with respect to the proposed actions of individual governments. The U.S. approach, instead, is an application of basic game theory—the prisoner's dilemma—whereby individual, similarly situated actors, in negotiations with a more powerful actor, are denied the opportunity to communicate with each other. Of course, U.S. allies are not precluded from coordinating with each other as they negotiate with NASA—indeed, coordination is their best strategy—but the bilateral context makes this more difficult.

We have been here before with different outcomes. In 1945, the United States claimed exclusive jurisdiction over the resources of the continental shelf located offshore its territory. Although novel, the claim quickly became part of international law, because its framing made the same right available to every coastal nation. Those nations responded by making the same claim themselves.

During negotiation of the 1982 UN Convention on the Law of the Sea (UNCLOS), the United States demanded that private companies have access to deep seabed resources beyond the continental shelf and be answerable to national regulators only. Most other nations wished the deep seabed to be recognized as "common heritage of mankind," with mining subject to international regulation and oversight. The latter view prevailed, mostly because of coordinated negotiating by developing nations. The United States has not ratified UNCLOS but accepts its provisions as customary international law.

With its proposed Artemis Accords, the United States is overlooking best practice with regard to the sustainable development of space. Instead of pressing ahead unilaterally and bilaterally, the United States should support negotiations on space mining within the UN Committee on the Peaceful Uses of Outer Space, the same multilateral body that drafted the five major space treaties of the 1960s and '70s. Meanwhile, NASA's actions must be seen for what they are—a concerted, strategic effort to redirect international space cooperation in favor of short-term U.S. commercial interests, with little regard for the risks involved. ■



Concept for robotic construction of buildings on the Moon. Lunar mining is envisaged to be necessary for developing such infrastructure.

ILLUSTRATION: NASA

REFERENCES AND NOTES

1. J. Foust, "NASA offers to buy lunar samples to set space resources precedent," *SpaceNews*, 10 September 2020; https://spacenews.com/nasa-offers-to-buy-lunar-samples-to-set-space-resources-precedent/.
2. International Space Exploration Coordination Group, "The global exploration roadmap" (Publication NP-2018-01-2502-HQ, NASA, 2018).
3. NASA, The Artemis Accords: Principles for a safe, peaceful, and prosperous future (2020); https://www.nasa.gov/specials/artemis-accords/index.html.
4. M. B. Duke, L. R. Gaddis, G. J. Taylor, H. H. Schmitt, *Rev. Mineral. Geochem.* **60**, 597 (2006).
5. E. Jarosewich, *Meteoritics* **24**, 323 (1990).
6. K. Lodders, H. Palme, H. P. Gail, in *Landolt-Börnstein—Group VI Astronomy and Astrophysics*, J. E. Trümper, Ed. (Springer, 2009), vol. 4B, chap. 4.4, pp. 560–630.
7. L. Schlüter, A. Cowley, *Planet. Space Sci.* **181**, 104753 (2020).
8. E. Ostrom, J. Burger, C. B. Field, R. B. Norgaard, D. Policansky, *Science* **284**, 278 (1999).
9. K. Shahar, D. Greenbaum, *New Astron.* **4**, 208 (2020).
10. National Academies of Sciences, Engineering, and Medicine, "Review and assessment of planetary protection policy development process" (Consensus Study Report, The National Academies Press, 2018).
11. P. T. Metzger, paper presented at the Impact of Lunar Dust on Human Exploration Workshop, Houston, TX, 11 to 13 February 2020.
12. L. Fladeland, A. C. Boley, M. Byers, paper presented at the First International Orbital Debris Conference, Sugar Land, TX, 9 to 12 December 2019.
13. M. Arakawa *et al.*, *Science* **368**, 67 (2020).
14. P. Wiegert, *Planet. Sci. J.* **1**, 3 (2020).
15. P. C. Stern, *Int. J. Commons* **5**, 213 (2011).

10.1126/science.abd3402

Published by AAAS

A man pauses to watch the sunset in Romania during a punishing heat wave in August 2019.

PHOTO: ANDREEA ALEXANDRU/AP PHOTO

BOOKS *et al.*

FICTION

Plotting a path to a more sustainable future

Persistent political pressure staves off global warming in a new climate allegory

By **Carl Abbott**

Renowned science fiction writer Kim Stanley Robinson has long been fascinated by climate-modifying technologies such as geoengineering and terraforming, but his passion lies in the necessity of political action to achieve a sustainable future. Based in an affectionately described Zurich, whose normalcy contrasts with global crisis, the titular Ministry for the Future described in Robinson's new book is a small agency within the United Nations with a mandate to encourage nations to realize the goals of the Paris Agreement on climate change.

The book follows Mary Murphy, a middle-aged Irish diplomat who heads the ministry from the mid-2020s through the 2040s. Her story intertwines with that of Frank May, who survives a heat wave that kills 20 million people in 2025 but never regains his mental stability. May seeks peace by living rough in Zurich and working with refugees, but he is eventually imprisoned for breaking into Murphy's apartment—an act driven by anguish over the slow pace of official action. She visits him in prison, and they slowly build an unlikely friendship. May serves as a stand-in for the damaged Earth that Murphy is trying to protect.

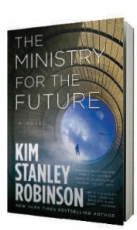
This is a book of ideas, not a plot-driven page-turner, which advances through 106 short chapters. Some are episodes of traditional action—activists take over a Davos meeting and hold the economic elite captive, Murphy escapes from assassins over an Alpine pass. Interspersed are loving portraits of landscapes and ecosystems, as well as capsule essays on climate science, ecology, politics, economic history, and global finance. Chapters that give first-person voice to a photon, a carbon atom, and a herd of caribou put human concerns in perspective. Readers familiar with Robinson's work will see parallels to his *Science in the Capital* trilogy (*1*–*3*), which followed fictional scientists and administrators at the National Science Foundation struggling to translate climate science into action.

Along with some of Robinson's favorite alternative technologies, such as dirigible airships and hybrid clipper ships powered by sails and solar energy, *The Ministry for the Future* explores two prominent science-based interventions that have been proposed as potential mitigators of global warming. Several chapters detail efforts to stabilize Antarctic and Greenland glaciers by drilling into their respective bases and pumping out the meltwater that lubricates their speedy advance. Several others imagine how early efforts to restore wildlife corridors might be scaled up to rewild much of the planet.

Ultimately, however, the book privileges global politics over geoengineering. For example, over several pivotal scenes, Murphy works to convince the heads of the world's largest central banks to back carbon coins—monetized credits awarded for keeping fossil fuels in the ground and for sequestering atmospheric carbon.

Climate action in Robinson's fictional future comes from strategy sessions in Zurich and from lobbying the Federal Reserve and the Chinese Ministry of Finance, but it also comes from local initiatives. One important chapter is simply a list of hundreds of grassroots organizations—from an Argentine permaculture initiative to a Zimbabwean ecological land trust—whose cumulative impact forces power brokers to consider alternatives to neoliberalism. Citing, both as an example and a caution, the revolution that swept through Europe in 1848, the novel envisions a moment when both governors and the governed realize that a new set of ideas has more power and relevance than the old.

Like the protagonists of Robinson's earlier books, Murphy is a hero because she is persistent. Her Ministry for the Future pushes, prods, and persuades decision-makers to reverse greenhouse gas accumulation, but even as she retires from her post, she knows that progress does not mean victory—just a measure of positive change. She has spent her career struggling with "the clot of trends, the Gordian knot of the world," tugging at this strand and that to loosen the snag. The future, she realizes, will be filled with new and different challenges. "We will keep going," she muses to herself, "because we never really come to the end." Like Murphy, we must all keep at it. ■



The Ministry for the Future
Kim Stanley Robinson
Orbit, 2020. 576 pp.

The reviewer is an emeritus professor in the Nohad A. Toulan School of Urban Studies and Planning, Portland State University, Portland, OR, USA. Email: abbottc@pdx.edu

REFERENCES AND NOTES

1. K. S. Robinson, *Forty Signs of Rain* (HarperCollins, 2004).
2. K. S. Robinson, *Fifty Degrees Below* (Spectra, 2005).
3. K. S. Robinson, *Sixty Days and Counting* (Spectra, 2007).

10.1126/science.abd9837

Published by AAAS

TECHNOLOGY

Moving beyond buzzwords

Rejecting conventional connotations, two books urge readers to rethink innovation

By **Dov Greenbaum**[1,2] *and* **Mark Gerstein**[2]

Two new books—Matt Ridley's *How Innovation Works* and Lee Vinsel and Andrew Russell's *The Innovation Delusion*—reflect on society's semantic satiation with "innovation." Both find that the term has been reduced to a buzzword; however, from this shared starting point, they take distinctly different directions.

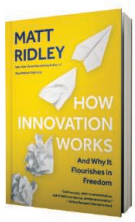
In *How Innovation Works*, Ridley aims to upend the notion that innovation emanates from a lone inventor. Although he finds innovation to be "the most important fact about the modern world," he reports that what is popularly perceived as "invention" is really an extended innovation process—an evolutionary, as opposed to a revolutionary, process.

Just as DNA recombination often results in more-successful traits than those that arise from random mutations, so too, he argues, do innovations occur when humans mingle and exchange ideas, whether in cultural hotbeds such as today's Silicon Valley or in the Renaissance's wealthy city-states. Ridley maintains, however, that even in these incubators, innovation is a bottom-up, iterative process that is often the result of aftermarket tinkering by many users rather than the actions of the initial inventors. To support his assertion, he meticulously documents various case studies of innovation in domains ranging from public health to communications.

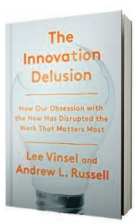
Invention may be the initial trigger for a new technology, acknowledges Ridley, but innovation happens as the initial idea evolves into marketable goods. To maximally benefit from science and technology, he argues that we must focus less on revolutionaries and more on the follow-on incremental tinkerers who substantially benefit from the unencumbered exchange of knowledge.

In stark contradiction to Ridley's efforts to bolster the concept of innovation and save it from its success, in their book, *The Innovation Delusion*, authors Lee Vinsel and Andrew Russell expose society's Faustian bargain with it. In their opinion, strategies to boost innovation, such as emphasizing STEM education—which, they argue, often advance "the interests of universities and corporations" rather than those of students—have led us to a misplaced focus on innovation for innovation's sake. Moreover, this misguided emphasis ignores what matters most in a thriving society: maintenance.

Vinsel and Russell note that they are "sick of hearing about what's good for Silicon Valley, and what the innovating classes think is good for us." However, as much as they dislike the innovation mind-set, many of the maintainers that Vinsel and Russell celebrate could be considered innovators. They cite John Day Jr., for example, who pioneered the idea that computers could be used to manage maintenance and advanced the idea that "every corrective maintenance action should be balanced by six preventive maintenance actions." Readers might also take Vinsel and Russell to task for not appreciating how innovation can supplant the need for maintenance—think cell phones versus phone booths.



How Innovation Works
Matt Ridley
HarperCollins, 2020. 416 pp.



The Innovation Delusion
Lee Vinsel and Andrew L. Russell
Currency, 2020. 272 pp.

Although they diverge on many issues, both books posit that innovation is not only the purview of Silicon Valley. Vinsel and Russell, for example, quote the economist Robert Gordon, who noted that computing technologies "do not hold a candle to the technological advances between 1870 and 1970," such as electricity, concrete, steel, automobiles, and airplanes. Quoting the venture capitalist Peter Thiel, Ridley expresses a similar sentiment: "We wanted flying cars; instead we got 140 characters."

Both books also cite government regulation as inhibitory to the innovation process. Ridley, for example, uses numerous examples to show how intellectual property (IP) laws inhibit innovation. In particular, he notes how expansive regulation impedes innovation in areas such as the genetic modification of foods by disincentivizing the trial and error necessary for such technologies to advance. IP regulations are also limitations on maintenance, according to Vinsel and Russell. The "right to repair" is an ongoing fight in Congress, hampered by restrictive IP licenses that limit owners' ability to fix property ranging from tractors to cell phones.

The titles of both books under review herein reveal that even those who aim to change our understanding of innovation are themselves unable to fully reject conventional definitions of the word. Were we consulted, we would have advised leaving "innovation" out of each title lest it perpetuate an outdated connotation. This criticism aside, both books are valuable in forcing us to look beyond uninformative buzzwords to appreciate what is truly valuable to society. ■

10.1126/science.abd9805

[1]Zvi Meitar Institute for Legal Implications of Emerging Technologies, Radzyner Law School, Interdisciplinary Center Herzliya, Herzliya, Israel. [2]Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. Email: dov.greenbaum@aya.yale.edu

A bridge collapse in Tempe, Arizona, highlights the importance of investment in infrastructure maintenance.

PHOTO: MPI34/MEDIAPUNCH/IPX/AP PHOTO

Published by AAAS

Feral horses in the Australian Alps threaten the environment, forcing policy-makers to weigh humane conservation practices against less costly approaches.

PHOTO: © LANGRISH, BOB/ANIMALS ANIMALS

Edited by **Jennifer Sills**

Retraction

The supplementary materials (SM) for our Report "Proton transport enabled by a field-induced metallic state in a semiconductor heterostructure" (*1*) contained mistakes. Fig. S10 was incorrect. The image had been previously saved in an experimental database. The figure should have shown a schematic illustration. This mistake occurred when computer data was transferred from one student to another. Fig. S9B also raised questions. The image appears to be a copy of figure 8B in a 2018 paper in the *International Journal of Hydrogen Energy* (*2*). The tests producing the data for these two figures were performed during the same period of time. Because of the similarity between the *IV-IP* measurements of SN-120 and the Na_xCoO_2/CeO_2 (x = 0.55) measurements, these data were wrongly saved in the same folder. Therefore, the final plotted fig. S9B in *Science* erroneously contains data from the figure in the 2018 paper. We repeated the experiment to replicate the data and found that the new outputs of two Na_xCoO_2/CeO_2 (x < 0.6) cells are above 550 mW cm^{-2} at 520°C, consistent with the data in fig. S9.

These mistakes were related to recent location and personnel changes. Between June and September 2019, all laboratories of the Faculty of Materials Science and Chemistry, including ours, had to move to a new campus 30 km away. During the same time, author L. Liu graduated, but the ordinary data transfer to successors was temporarily discontinued. The manuscript was finalized between mid-December 2019 and March 2020, during the unprecedented lockdown of the city of Wuhan for coronavirus disease 2019. All authors who prepared the main manuscript and SM data were isolated in different cities and countries, without access to lab computers or database resources, which led to miscommunications.

We take full responsibility for the accidental mistakes in the SM. We stand by our experimental design, theoretical calculations, main data, and analysis of low-temperature proton ceramic fuel cells. However, given the mistakes in the SM, we have decided to retract the Report. All authors except for B. Zhu, M. Huang, M. Akbar, and J. S. Kim approved this Retraction.

Y. Wu[1], L. Liu[1], Q. Shi[1], C. Chen[2], J. Wei[3], J. F. Li[3], L. R. Zheng[4], H. B. Song[1]*

[1]Engineering Research Center of Nano-Geo Materials of Ministry of Education, Faculty of Materials Science and Chemistry, China University of Geosciences, Wuhan, 430074, China. [2]Huazhong University of Science and Technology, Wuhan, 430074, China. [3]College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361005, China. [4]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, 100049, China.
*Corresponding author. Email: songhb@cug.edu.cn

REFERENCES AND NOTES

1. Y. Wu, B. Zhu, M. Huang, L. Liu, Q. Shi, M. Akbar, C. Chen, J. Wei, J. F. Li, L. R. Zheng, J. S. Kim, H. B. Song, *Science* **369**, 184 (2020).
2. L. Liu *et al.*, *Int. J. Hydrogen Energ.* **43**, 12739 (2018).

10.1126/science.abe7205

Hidden ethical costs of conservation

We agree with N. Sekar and D. Shillers ("Engage with animal welfare in conservation," Policy Forum, 7 August, p. 629) that welfare considerations are an important part of the myriad values in conservation. However, more humane conservation tactics are often costlier. The money and resources spent on humane conservation practices detract from the resources available for other conservation projects. This hidden ethical cost of humane conservation must be considered when proposing new approaches to conservation management.

Biodiversity conservation is associated with a variety of values that often vary implicitly between individuals and organizations (*1*, *2*). These values are rarely noticed until they conflict, such as when invasive-species culling is proposed. Compassionate conservationists have responded to this challenge with humane but more expensive animal management alternatives including fertility control and trapping and rehoming animals (*3*). Feral horse management in the Australian Alps with passive trapping costs approximately AU$1116 per horse, whereas aerial culling costs approximately AU$85 per horse (*4*). For the same cost, more culling could be carried out, resulting in a greater net area of biodiversity conserved. Alternatively, savings achieved by implementing aerial culling could be spent on

Published by AAAS

other conservation projects in the region. In addition to promoting biodiversity conservation, these projects could benefit the welfare of other animals by protecting threatened native species (*5*).

Considerations of what one could have done otherwise with a given resource is a foundational principle of the effective altruism movement, an evidence-based ethical framework (*6*). Emerging applications of this framework to conservation include considerations of animal welfare and offer promising solutions to complex cases of value conflict (*7*, *8*).

Kate E. Lynch[1]* and Daniel T. Blumstein[2]
[1]Department of Philosophy, The University of Sydney, Camperdown, NSW 2006, Australia. [2]Department of Ecology and Evolutionary Biology, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA 90095, USA.
*Corresponding author.
Email: kate.lynch@sydney.edu.au

REFERENCES AND NOTES

1. C. Gamborg *et al.*, *Nat. Educ. Knowl.* **3**, 8 (2012).
2. J. A. Newman *et al.*, *Defending Biodiversity: Environmental Science and Ethics* (Cambridge University Press, Cambridge, UK, 2017).
3. A. Wallach *et al.*, *Conserv. Biol.* **32**, 1255 (2018).
4. N. J. Beeton, C. N. Johnson, *Ecol. Manage. Restor.* **20**, 57 (2019).
5. D. A. Driscoll *et al.*, *Ecol. Manage. Restor.* **20**, 63 (2019).
6. W. McAskill, *Doing Good Better: How Effective Altruism Can Help You Make a Difference* (Penguin Random House, 2015).
7. K. E. Lynch, D. T. Blumstein, *TREE* **35**, 857 (2020).
8. B. S. Freeling, S. D. Connell, *TREE* **35**, 3 (2020).

10.1126/science.abe2505

Pragmatic animal welfare is independent of feelings

In their Policy Forum "Engage with animal welfare in conservation" (7 August, p. 629), N. Sekar and D. Shiller state that the "overwhelming evidence that animals think and feel" is the basis for their call to include animal welfare in conservation practices. This feelings-based approach is problematic because there is substantial scientific uncertainty about whether taxa such as fish are sentient and, therefore, able to feel pain and suffer (*1*, *2*).

In recreational fishing, animal welfare concepts are embedded in global international conservation policies (*3*, *4*) and in local welfare actions (*5*), despite the uncertainty about fish sentience. These activities are motivated by the reality that fish populations are composed of individuals whose well-being is important to the conservation of populations and fisheries, regardless of whether the animal is able to think and feel. Moreover, many users of fish respect the life, function, and welfare of individual fish and act accordingly, independent of whether they think that the animal can feel pain (*5*, *6*).

Recreational fishing policies recognize the importance of animal welfare regardless of whether fish are sentient.

An effective application of animal welfare in conservation is possible—and is perhaps more effective and convincing to stakeholders—without invoking or relying on concepts such as consciousness, sentience, or pain (*5*, *7*). A pragmatic approach to animal welfare that relies on objective and measurable endpoints of animal well-being is more likely to gain support among stakeholders and be implemented in practice than a feelings-based framework that is based on concepts that are difficult to define and cannot be readily measured in many taxa (*5*, *7*).

Robert Arlinghaus[1,2]*, Ian G. Cowx[3], Brian Key[4], Ben K. Diggles[5], Alexander Schwab[6], Steven J. Cooke[7], Anne Berit Skiftesvik[8], Howard I. Browman[8]
[1]Department of Biology and Ecology of Fishes, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [2]Division of Integrative Fisheries Management, Faculty of Life Sciences, Humboldt-Universität zu Berlin, Berlin, Germany. [3]University of Hull International Fisheries Institute, Hull HU6 7RX, UK. [4]School of Biomedical Sciences, University of Queensland, Brisbane, QLD 4072, Australia. [5]DigsFish Services, Banksia Beach, QLD 4507, Australia. [6]Eichelmändli Verlag, Hofstetten SO, Switzerland. [7]Fish Ecology and Conservation Physiology Laboratory, Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada. [8]Institute of Marine Research, Ecosystem Acoustics Group, Austevoll Research Station, 5392 Storebø, Norway.
*Corresponding author.
Email: arlinghaus@igb-berlin.de

REFERENCES AND NOTES

1. J. D. Rose *et al.*, *Fish Fish.* **15**, 97 (2014).
2. B. Key, *Biol. Phil.* **30**, 149 (2015).
3. Food and Agriculture Organization of the United Nations (FAO), "FAO technical guidelines for responsible fisheries recreational fisheries" (2012).
4. European Inland Fisheries Advisory Commission (EIFAC), "EIFAC code of practice for recreational fisheries" (FAO, 2008).
5. R. Arlinghaus *et al.*, *J. Fish Biol.* **75**, 2448 (2009).
6. J. C. Evans, *With Respect for Nature: Living as Part of the Natural World* (Sunypress, 2015).
7. M. S. Dawkins, *J. Zool.* **301**, 1 (2017).

COMPETING INTERESTS

This piece reflects the views of the authors and not those of their organizations or relevant funding sources.

10.1126/science.abe3397

Animal welfare science aids conservation

In their Policy Forum "Engage with animal welfare in conservation" (7 August, p. 629), N. Sekar and D. Shiller overlooked a long history of conservation-related animal welfare science. The integration of animal welfare science and conservation spans at least 60 years (*1*) and has been applied to a broad range of wildlife management activities (*2*) and interdisciplinary research (*3*–*5*). Understanding and incorporating animal welfare science can benefit conservation efforts.

Animal welfare science is not synonymous with opposition to intentional killing of wildlife (compassionate conservation) (*6*, *7*). Rather, animal welfare science uses quantitative measurements to assess harmful and positive impacts of human activities on animals (*8*). Traditionally, the harms are weighed against conservation benefits to justify (or rule out) a management action. Approaches such as compassionate conservation may, perhaps counterintuitively, worsen animal welfare outcomes and make biodiversity conservation more difficult (*9*).

Sekar and Shiller use prescriptive advocacy framing that does not represent animal welfare science. Stipulating that conservation agencies should avoid factory farming products does not reflect scientific quantification and comparison of harms posed by this and other human activities. Processes such as land clearing (*10*) may pose greater animal welfare impacts when all wild sentient species and types of harm (*11*) are considered.

We agree that increased animal welfare focus is warranted in conservation. Progress will be expedited by wider collaboration with animal welfare scientists. Decades-old

PHOTO: FLORIAN MÖLLERS

Published by AAAS

scientific tradition in this space teaches us to prioritize objective assessment of harms rather than deferring to advocacy.

Jordan O. Hampton[1]*, Sandra E. Baker[2], Ngaio J. Beausoleil[3], Marc Cattet[4], David M. Forsyth[5,6], Clive R. McMahon[7], Gilbert Proulx[8], Bruce Warburton[9]
[1]School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA 6150, Australia. [2]Wildlife Conservation Research Unit, Department of Zoology, Recanati-Kaplan Centre, University of Oxford, OX13 5JL, UK. [3]Animal Welfare Science and Bioethics Centre, School of Veterinary Science, Massey University, Palmerston North 4442, New Zealand. [4]RGL Recovery Wildlife Health and Veterinary Services, Saskatoon, SK S7H 4A6, Canada. [5]Vertebrate Pest Research Unit, Department of Primary Industries, Orange, NSW 2800, Australia. [6]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia. [7]IMOS Animal Satellite Tagging, Sydney Institute of Marine Science, Mosman, NSW 2088, Australia. [8]Alpha Wildlife Research & Management Ltd., Sherwood Park, AB T8H 1W3, Canada. [9]Landcare Research, Lincoln 7640, New Zealand.
*Corresponding author. Email: jordan.hampton@murdoch.edu.au

REFERENCES AND NOTES

1. B. Warburton, J. Hall, *New Zealand J. Zool.* **22**, 39 (1995).
2. S. E. Baker *et al.*, *BioScience* **63**, 928 (2013).
3. D. Fraser, *Anim. Welfare* **19**, 121 (2010).
4. B. Warburton, B. Norton, *J. Wildl. Manag.* **73**, 158 (2009).
5. A. M. Harvey *et al.*, *Animals* **10**, 148 (2020).
6. A. D. Wallach *et al.*, *Conserv. Biol.* **32**, 1255 (2018).
7. N. J. Beausoleil, *Animals* **10**, 257 (2020).
8. C. R. McMahon *et al.*, *Wildl. Res.* **39**, 375 (2012).
9. A. Callen *et al.*, *Biol. Conserv.* **241**, 108365 (2020).
10. H. Finn, N. Stephens, *Wildl. Res.* **44**, 377 (2017).
11. D. Fraser, A. M. MacRae, *Anim. Welfare* **20**, 581 (2011).

10.1126/science.abe2171

Response

We agree with Lynch and Blumstein that decision-makers must consider the opportunity costs—financial and ethical—of conservation programming. Making decisions in the face of such trade-offs will inevitably require value judgments. Decision-makers will likely continue to conclude that interventions that harm individual animals, such as culling, are sometimes merited by the benefits and necessitated by financial constraints.

Even so, explicit consideration of animal welfare could lead to affordable improvements when ideal alternatives are unworkable. For instance, evidence suggests that aerial culling operations might entail acceptable levels of duress for feral horses if done properly (*1*). For a marginal increase in cost, conservation organizations could promote oversight by independent animal welfare experts to ensure adherence to best practices during unavoidable culling operations.

Focusing too much on the current financial costs of humane alternatives could cause the conservation community to discount the value of prioritizing animal welfare. We hope that recognition of the moral relevance of individual animals' life experiences could lead to innovations that circumvent difficult trade-offs. For instance, conservationists have already demonstrated the promise of contraceptives to address population control for horses (*2*). Giving animal welfare the appropriate weight could motivate efforts to reduce the costs of contraceptive programs. Prioritization drives innovation.

Arlinghaus *et al.* are likely correct that an instrumental approach, which sets aside the issue of suffering, eases the introduction of some safeguards for animal welfare in recreational fishing. However, experts remain divided on the capacity for consciousness in fish (*3*, *4*). If fish can suffer, protections based solely on instrumental considerations would be inadequate. Conservation organizations should, wherever practical, establish policies that reflect the diversity of evidence on suffering in fish. The uncertainty about what fish can experience—in contrast to the expert consensus about the capacity for mammals and birds to suffer—highlights how animal welfare institutions in conservation must navigate diverse challenges across taxa. Formal concern for animal welfare could not only shape conservation practice but also influence the direction of fundamental research on the quality of animal life.

Hampton *et al.* describe the work done by animal welfare scientists in conservation. Our Policy Forum did not claim that animal welfare science in conservation would be novel but rather that it deserves more attention and institutional support. The public looks to conservation organizations for leadership in how to ethically engage with other species. As such, institutional safeguards for animal welfare should be at least as much a priority in conservation as in academic research. The small minority of conservation organizations that do explicitly recognize animal welfare concerns [e.g., (*5*, *6*)] have demonstrated that policies promoting animal welfare in conservation are practicable.

We agree with Hampton *et al.* that animal welfare science should not be biased by unsubstantiated pre-existing beliefs. However, once enough evidence has accumulated, evidence-driven advocacy can be crucial for translating science into societal improvement. The harms from factory farming to billions of domestic animals [e.g., (*7*, *8*)], the climate, and biodiversity (*9*, *10*) merit decisive action from conservation organizations. Whereas some substitutes for factory farming might have negative biodiversity consequences, other alternatives—such as diets that largely exclude animal products—can offer a clear net win for both animal welfare and biodiversity (*9*, *10*).

Conservation organizations frequently engage in advocacy to build a better planet for animals, including efforts to shape the cultural and consumptive practices of people around the world (*11*, *12*). When presented with evidence-based opportunities to improve the quality of animal life, conservation organizations must be receptive to altering their own practices as well.

Nitin Sekar[1]* and Derek Shiller[2]
[1]Wildlife and Habitats Division, WWF-India, New Delhi, India. [2]The Humane League, New York, NY 10022, USA.
*Corresponding author.
Email: nitin.sekar@gmail.com

REFERENCES AND NOTES

1. J. O. Hampton *et al.*, *Wildl. Res.* **44**, 97 (2017).
2. C. M. V. Nuñez *et al.*, *Conserv. Physiol.* **5**, 1 (2017).
3. V. Braithwaite, *Do fish feel pain?* (OUP, Oxford, 2010).
4. J. D. Rose *et al.*, *Fish Fish.* **15**, 97 (2014).
5. International Fund for Animal Welfare, Guiding Principles (2020); www.ifaw.org/about/guiding-principles.
6. Sierra Club Board of Directors, "Policy on trapping of wildlife" (2012); www.sierraclub.org/sites/www.sierraclub.org/files/uploads-wysiwig/Trapping-Wildlife.pdf.
7. R. J. Julian, *Poult. Sci.* **77**, 1773 (1998).
8. M. R. N. Bruijnis *et al.*, *Animal* **7**, 167 (2013).
9. J. Poore, T. Nemecek, *Science* **360**, 987 (2018).
10. B. Machovina *et al.*, *Sci. Total Environ.* **536**, 419 (2015).
11. Global Wildlife Conservation, #endthetrade (2020); https://endthetrade.com/.
12. R. Fobar, "Shark fin is banned in 12 U.S. states—but it's still on the menu," *National Geographic* (2019).

COMPETING INTERESTS

This piece reflects the views of the authors and not the official positions of their organizations.

10.1126/science.abe5111

TECHNICAL COMMENT ABSTRACTS

Comment on "Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates"

Michael J. Braun, Gerald S. Wilkinson, Brian S. Cade

Chen and Pfennig (Reports, 20 March 2020, p. 1377) analyze the fitness consequences of hybridization in toads but do not account for differences in survival among progeny. Apparent fitness effects depend on families with anomalously low survival, yet survival is crucial to evolutionary fitness. This and other analytical shortcomings demonstrate that a conclusion of adaptive mate choice is not yet justified.

Full text: dx.doi.org/10.1126/science.abd3905

Response to Comment on "Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates"

Catherine Chen and Karin S. Pfennig

Braun *et al.* contend that we did not account for survival, but we did. Differential survival does not alter our conclusions, which were also robust to removing anomalous families. They ignore the study system's natural history justifying our fitness measures, while failing to account for our behavioral data. We stand by our conclusion that females adaptively choose among heterospecific males.

Full text: dx.doi.org/10.1126/science.abd5373

Published by AAAS

Cite as: M. J. Braun *et al.*, *Science* 10.1126/science.abd3905 (2020).

Comment on "Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates"

Michael J. Braun[1,2]*, Gerald S. Wilkinson[2]*, Brian S. Cade[3]*

[1]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA. [2]Department of Biology, University of Maryland, College Park, MD 20742, USA. [3]U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO 80526, USA.

***Corresponding author. Email: braunm@si.edu (M.J.B.); wilkinso@umd.edu (G.S.W.); cadeb@usgs.gov (B.S.C.)**

Chen and Pfennig (Reports, 20 March 2020, p. 1377) analyze the fitness consequences of hybridization in toads but do not account for differences in survival among progeny. Apparent fitness effects depend on families with anomalously low survival, yet survival is crucial to evolutionary fitness. This and other analytical shortcomings demonstrate that a conclusion of adaptive mate choice is not yet justified.

Chen and Pfennig (*1*) extend the potential impact of hybrid zone studies with a report of adaptive heterospecific mate choice in spadefoot toads, a result with far-reaching implications (*2*). However, survival is crucial to evolutionary fitness (*3*), and hybrid vigor, while common, does not always predict fitness. We wondered how consideration of these issues would affect their conclusions.

Central to their claim is an experiment in which 20 female *Spea bombifrons* were paired with 20 male *Spea multiplicata*. Survival and growth of up to 50 progeny per pair were monitored for 12 days, when several measures were taken [mass, snout-vent length (SVL), Gosner stage (an ordinal scale of development)]. These measures were highly correlated, so the authors used principal components analysis to derive an overall proxy of tadpole fitness (PC1). They then analyzed several linear models to explore whether variation in this proxy was correlated with parental size, condition, genotype, and sire vocalizations. This analysis resulted in their figure 1, which illustrates their conclusion that tadpole fitness is predicted by father's call (sire pulse rate).

We noticed that their results might depend on three outlier families at the upper or lower ends of the observed range in sire pulse rate (their figure 1). These three families also had low survival rates (≤30% versus mean of 64% for 17 families with surviving tadpoles) [table S2 of (*1*)]. The authors concluded that "Hybrid tadpole survivorship was not significantly predicted by any parental traits" [supplementary text of (*1*)], but they only considered linear models. In fact, a nonlinear, unimodal relationship between survival and sire pulse rate better fits the data (Fig. 1A), which suggests that survival should not be ignored. Furthermore, three families with zero survival were excluded (*4*) and two of these also had low sire pulse rates (Fig. 1A). The distribution of PC1 is surely truncated in families with low survival, but it is difficult to assess the impact of truncation given the experimental design.

We were also concerned about relying on measures of tadpole growth as the sole fitness metric in a hybrid system. Hybridization is often maladaptive, as it is for female *S. multiplicata* in this system (*2*, *5*). Heterosis (hybrid vigor) is common but does not always predict reproductive success: Mules are stronger than horses, but their lifetime fitness is zero. Although mass, SVL, and Gosner stage may sometimes be reasonable fitness proxies, these measures could reflect heterotic effects in this hybrid system. Moreover, these measures ignore critical later stages of the life cycle—including metamorphosis, survival, fertility, sex ratio, and mating success—all of which affect how selection operates on adult traits, such as male calling rate and female preference. Complete fitness estimates require following the fate of offspring from zygote until reproduction and ideally until death. These caveats, together with the question of how to deal with survival, led us to examine more closely the evidence that sire pulse rate predicts tadpole fitness.

Because the authors sought to estimate the partial effect of each fixed effect on mean PC1 scores, it was incorrect to model-average regression coefficients (*6*) without accounting for multicollinearity among the predictors in the standardization (*7*). In fact, neither is necessary. We explored the partial effects of sire pulse rate by examining estimates across the five models that included this term [table S3 of (*1*)] without standardization or model averaging. The models included random effects for each family and fixed effects for sire pulse rate and other parental predictors, each of which had a single unique value for each family. Thus, their analysis inherently considers parental predictors to enter

models both as categorical variables (through random effects on family) and as fixed linear effects.

Comparing a model with only family effects (equivalent to a categorical model for sire pulse rate, conditional R^2 = 0.289) to models that included a linear fixed effect for sire pulse rate indicated substantial lack of fit for the linear predictor (Fig. 1B) across the five models (marginal R^2 = 0.042 to 0.057). This analysis revealed that there is no consistent decline in mean PC1 scores up to a sire pulse rate of 27; in fact, most of the decline is due to a single family, which also had low survival, with the highest pulse rate. Estimating models excluding this family, we found that regression coefficients for sire pulse rate diminished greatly and were not statistically different from zero ($P > 0.05$) for three of the five models. The combination of lack of fit of the linear relationship of sire pulse rate, the low proportion of variance attributed to it, and small effect size of sire pulse rate ($\hat{\beta}$ = −1.214 to −0.928) relative to the estimated standard deviation ($\hat{\sigma}$ = 2.526 to 2.529) of the regressions indicates extensive overlap in PC1 distributions with minimal explanatory ability for sire pulse rate (Fig. 1C) (*8*, *9*).

Lastly, using a principal component score to estimate fitness is undesirable because it is a synthetic mathematical construct that lacks straightforward interpretation. In this case, the authors combine two continuous variables (mass, SVL) with an ordinal one (Gosner stage) using covariances that fail to account for scale differences. A more direct, easily interpreted measure of tadpole condition is to estimate mass (M) as an allometric function of size (SVL); that is, $M = \beta_0 SVL^{\beta_1}$ (*10*, *11*). We estimated this model for median mass at SVL with separate intercepts and exponents for each of 16 families using quantile regression (Fig. 2A) (*10*). We did not include Gosner stage as a predictor, because it had a weak effect (P = 0.079) and only reduced the coefficient of determination by 0.002. Estimated median mass at SVL = 14 mm with 95% confidence intervals for each family reveals considerable overlap and no distinct pattern when plotted against sire pulse rate (Fig. 2B) or tadpole survival (Fig. 2C); these findings indicate that tadpole condition was not related to either variable (also true for female SVL, not shown).

We conclude that the authors' claim of adaptive heterospecific mate choice is not well supported by their analyses or available data, but we encourage them to continue their insightful work on this remarkable system.

REFERENCES AND NOTES

1. C. Chen, K. S. Pfennig, Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates. *Science* **367**, 1377–1379 (2020). doi:10.1126/science.aaz5109 Medline
2. M. Zuk, Shallow ponds prompt fitness-favorable species interbreeding. *Science* **367**, 1304–1305 (2020). doi:10.1126/science.abb2398 Medline
3. C. R. Darwin, A. R. Wallace, On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zool. J. Linn. Soc.* **3**, 45–62 (1858).
4. Data on these three families were provided by the authors (C. Chen and K. S. Pfennig). One of these families produced no eggs; the other two produced eggs, but embryos failed to develop beyond the hatching stage.
5. L. K. Wünsch, K. S. Pfennig, Failed sperm development as a reproductive isolating barrier between species. *Evol. Dev.* **15**, 458–465 (2013). doi:10.1111/ede.12054 Medline
6. B. S. Cade, Model averaging and muddled multimodel inferences. *Ecology* **96**, 2370–2382 (2015). doi:10.1890/14-1639.1 Medline
7. J. Bring, How to standardize regression coefficients. *Am. Stat.* **48**, 209–213 (1994).
8. R. Xie, P. I. Nelson, Separation among distributions related by linear regression. *Am. Stat.* **57**, 33–36 (2003). doi:10.1198/0003130031117
9. A. Farmer, B. S. Cade, J. Torres-Dowdall, Fundamental limits to the accuracy of deuterium isotopes for identifying the spatial origin of migratory animals. *Oecologia* **158**, 183–192 (2008). doi:10.1007/s00442-008-1143-6 Medline
10. B. S. Cade, J. W. Terrell, B. C. Neely, Estimating geographic variation in allometric growth and body condition of blue suckers with quantile regression. *Trans. Am. Fish. Soc.* **140**, 1657–1669 (2011). doi:10.1080/00028487.2011.641885
11. J. Lleonart, J. Salat, G. J. Torres, Removing allometric effects of body size in morphological analysis. *J. Theor. Biol.* **205**, 85–93 (2000). doi:10.1006/jtbi.2000.2043 Medline
12. R scripts for all modeling in this Comment are deposited in the DRUM archive at the University of Maryland with the following identifier: http://hdl.handle.net/1903/26347.

ACKNOWLEDGMENTS

R. C. Bell and P. M. Dixon provided helpful comments on an earlier draft of this work.

19 June 2020; accepted 10 September 2020
Published online 9 October 2020
10.1126/science.abd3905

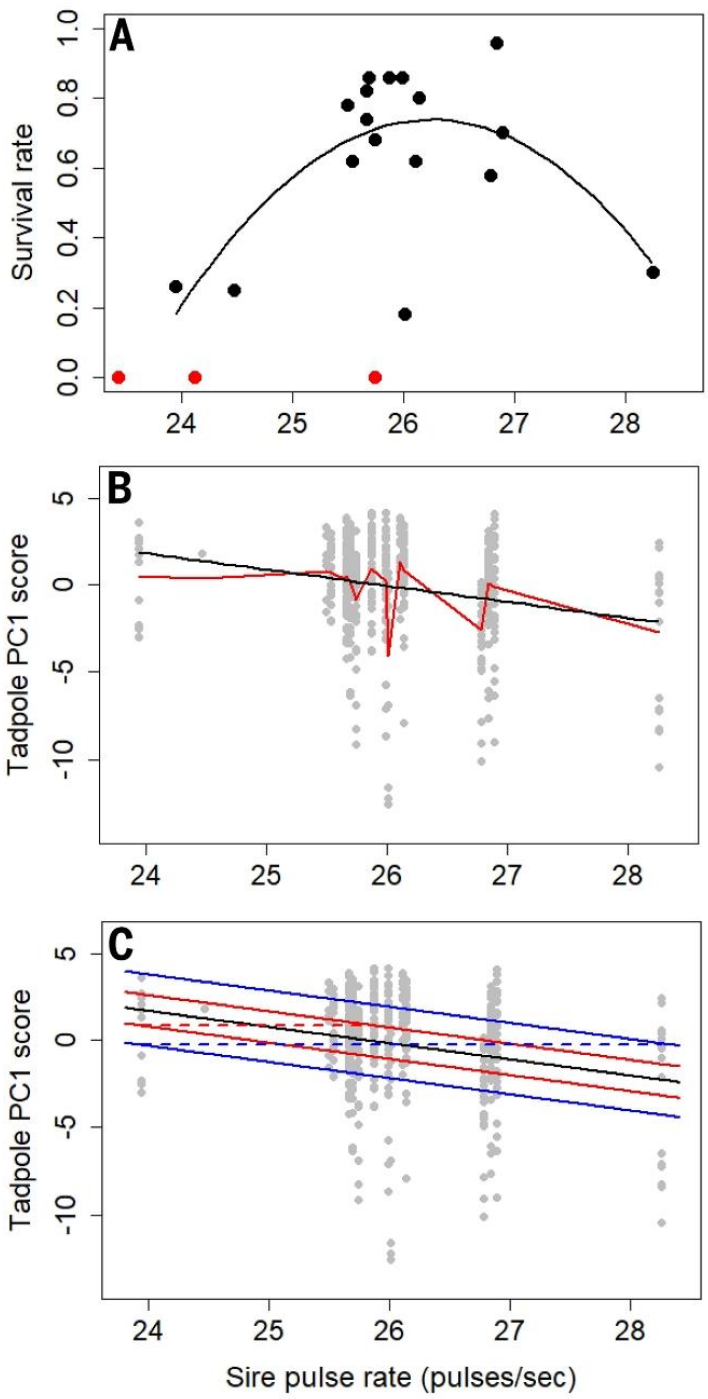


Fig. 1. Relationships between sire pulse rate and fitness proxies do not support adaptive mate choice. (**A**) Tadpole survival and sire pulse rates for 17 families analyzed by Chen and Pfennig (*1*) (black dots) and for three families not included in their analyses (red dots) (*4*). A nonlinear unimodal model (black line) for n = 17 families had P = 0.014 for the joint effect of the linear and quadratic terms. This model fits significantly better (ΔAICc = 6.35) than a linear model (*12*). AICc is the Akaike information criterion with correction for small sample sizes. (**B**) Estimated relationship of mean principal component 1 (PC1) scores and sire pulse rate for 17 families (n = 479 tadpoles) from a mixed-effects model with a random effect for family plus a linear fixed effect for sire pulse rate (black line) and for a model using only a random effect for family (red lines), which demonstrates lack of fit of the linear relationship. (**C**) Quantile splits for the separation of PC1 score distributions for tadpoles illustrate the lack of predictive power of sire pulse rate. The solid black line is the mean regression estimate from the mixed-effects model of Chen and Pfennig, as in (B). The lower and upper red lines are the 0.36 and 0.64 quantiles, respectively, and the lower and upper blue lines are the 0.21 and 0.79 quantiles, respectively, from the normal distribution. These correspond to quantile splits QS(1.93) = 0.36 and QS(4.31) = 0.21, indicating 36% and 21% overlap in PC1 probability distributions, respectively. The horizontal dashed red and blue lines depict separation distances = 1.93 (median − minimum) and 4.31 pulses/s (maximum − minimum), respectively. Quantile splits for the PC1 score distribution at any separation distance are computed as QS(separation distance) = $\Phi[-|\text{estimated slope for sire pulse rate}| \times (\text{separation distance})/(2 \times \text{estimated standard deviation of regression})]$, where Φ is the normal distribution function (*8*, *9*). Here, the estimated slope is −0.928 and the estimated standard deviation of the regression is 2.528.

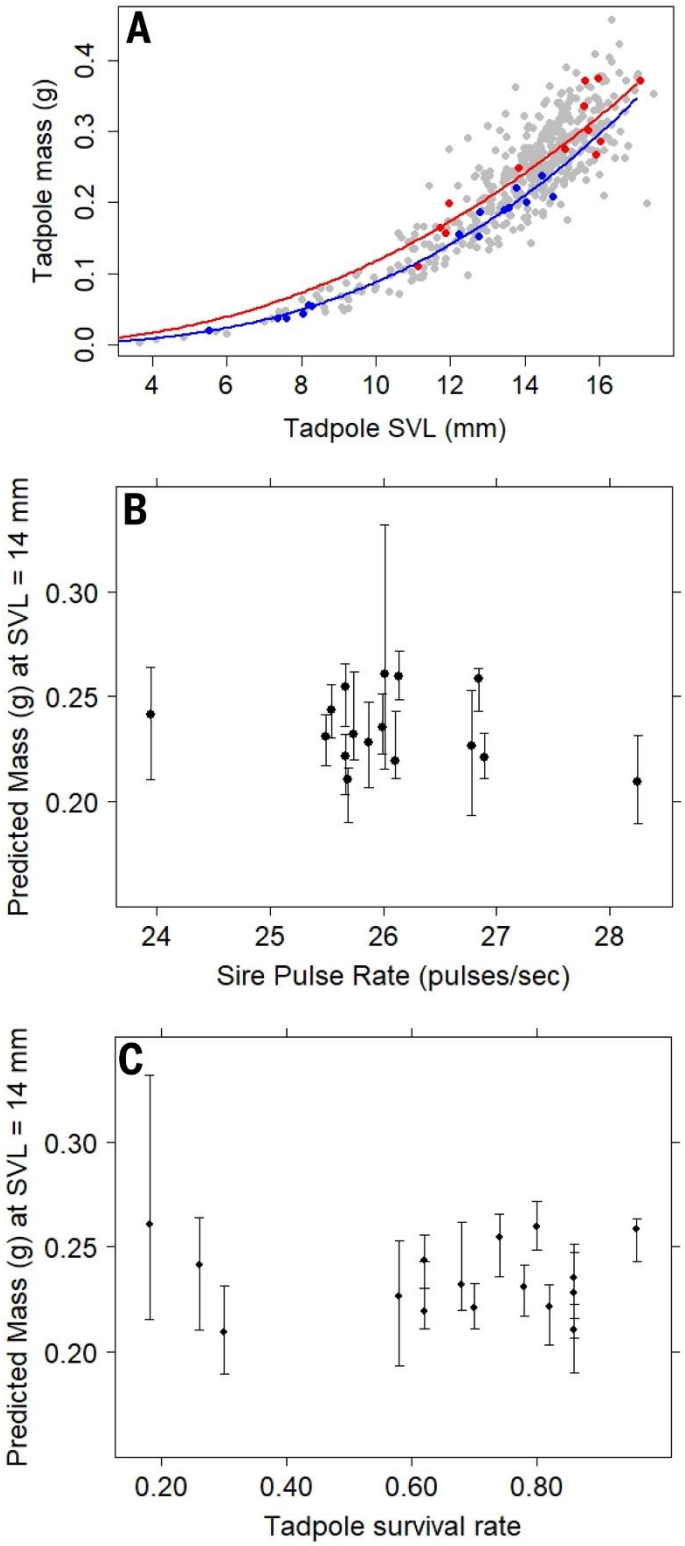


Fig. 2. Allometric modeling reveals extensive overlap in tadpole condition. (**A**) Allometric relationships of tadpole mass (g) and snout-vent length (SVL, mm) for 16 families of tadpoles from Chen and Pfennig (*1*). Family 15 with a single surviving tadpole was excluded from this analysis. This model indicates higher condition for tadpoles with greater mass at a common SVL achieved during the 12-day time interval. Estimated median regression relationships for the families with lowest (red) and highest (blue) sire pulse rates from the quantile regression model allow separate intercepts and slopes for each family, $p = 32$ parameters and $n = 478$. The allometric relationship was estimated in a linear model as $\log(M) = \log(\beta_0) + \beta_1 \log(SVL)$ and back-transformed to the multiplicative form. Here, we used $\log_{10}$ and included indicator variables for 16 families and their interaction with SVL. The coefficient of determination for this model is 0.691 in absolute deviations ($P < 0.001$), which is equivalent to $0.905 = 1.0 - (1.0 - 0.691)^2$ in squared deviations similar to R^2. (**B**) Median mass at SVL = 14 mm and 95% confidence intervals (bootstrapped) estimated for each of the 16 families and graphed with respect to their corresponding sire pulse rates. Similar patterns were obtained for SVL = 12 and 16 mm, not shown. (**C**) Same as (B) but now plotted against tadpole survival.

Cite as: C. Chen, K. S. Pfennig, *Science* 10.1126/science.abd5373 (2020).

Response to Comment on "Females engaging in adaptive hybridization prefer high-quality heterospecifics as mates"

Catherine Chen* and Karin S. Pfennig*

Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA.

***Corresponding author. Email: catherine.chen@unc.edu (C.C.); kpfennig@unc.edu (K.S.P.)**

Braun *et al.* contend that we did not account for survival, but we did. Differential survival does not alter our conclusions, which were also robust to removing anomalous families. They ignore the study system's natural history justifying our fitness measures, while failing to account for our behavioral data. We stand by our conclusion that females adaptively choose among heterospecific males.

Hybridization is adaptive if its fitness benefits outweigh its costs (*1*). This occurs for spadefoot toads, *Spea bombifrons*. Hybridization with *S. multiplicata* produces offspring with low fertility (males are sterile; females are less fecund). However, under some circumstances, hybrid tadpoles are more likely to survive than pure *S. bombifrons* tadpoles (*1*, *2*), and *S. bombifrons* females prefer to mate with heterospecifics when hybridization has this net fitness benefit (*1*) (Fig. 1). We examined whether *S. bombifrons* females further optimize fitness of their hybrid offspring in this crucial tadpole stage by preferring *S. multiplicata* males that sire higher-fitness tadpoles versus mating randomly with *S. multiplicata*. We found that the pulse rate of *S. multiplicata* male calls predicted hybrid tadpole fitness (*3*). Critically, *S. bombifrons* females preferred pulse rates of *S. multiplicata* males corresponding with fitter offspring only when hybridization was favored (*1*–*3*).

Braun *et al.* (*4*) contend that our conclusions are unsupported because we did not identify a quadratic relationship between survival and male pulse rate [we investigated a linear effect and found none (*5*)]. However, they do not account for parent size and condition (*3*), which should be included whenever evaluating relationships between fathers' sexual signals and offspring fitness (*6*, *7*). Using a logistic regression with tadpole survival/death as the response and family as a random effect reveals that the top three models include a quadratic effect of pulse rate (ΔAICc $\leq$ 0.98), and the fourth-best includes male condition only (ΔAICc = 1.70) [the fifth-best was the null model (ΔAICc = 4.58)] (*8*). Given these small AICc differences [<2 (*9*)], a quadratic relationship between pulse rate and offspring survival is suggested but not definitive.

Braun *et al.* also do not articulate how a quadratic relationship between survival and male pulse rate undercuts our findings. Differential survival weakens our conclusions if the pattern of survival opposes that of growth (i.e., if better-growing tadpoles are more likely to die). However, tadpoles were more likely to survive when from families with better average growth [using our original fitness metric, PC1 from a principal components analysis of snout-vent length (SVL), Gosner stage (GS), and mass]: logistic regression with family as a random effect; likelihood ratio test, $\chi^2(1) = 5.52$, $P = 0.019$; the same holds when separately examining SVL [$\chi^2(1) = 4.99$, $P = 0.025$] and GS [$\chi^2(1) = 6.03$, $P = 0.014$] [mass, which is confounded with SVL: $\chi^2(1) = 3.35$, $P = 0.067$]. Thus, our original analyses appear conservative.

Braun *et al.* do not account for our mate choice data. In two-choice tests, *S. bombifrons* females preferred heterospecific calls associated with higher tadpole fitness, and females only expressed those preferences when hybridization was favored (*1*–*3*). A curvilinear relationship between sire signal and offspring fitness would still demonstrate that male sexual signals predict hybrid offspring fitness, and females would still be expected to prefer heterospecific mates that sire high-fitness offspring. Specifically, if the relationship between male calls and offspring fitness were curvilinear, females should disfavor extreme pulse rates relative to those we used (*6*). Such "preference function" data (*6*) would provide a nuanced perspective on sexual selection exerted by *S. bombifrons* females on *S. multiplicata* males, but would not undercut our study.

Braun *et al.* also take issue with our measures of tadpole growth "as the sole fitness metric in a hybrid system" and argue that growth reflects heterosis (i.e., enhanced fitness of hybrids relative to both pure species). However, Braun *et al.* do not acknowledge crucial details motivating our study.

The spadefoot natural history and known fitness consequences across life stages, which we described (*3*), justifies our focus on tadpole growth (see also above). Spadefoot tadpoles develop in desert ponds (Fig. 1). Growth (in terms of size) and development time (measured using GS) have life-or-death consequences (Fig. 1C). Owing to intermediate—not heterotic—growth rates of hybrid tadpoles, hybrids have higher fitness than *S. bombifrons* (but not *S. multiplicata*) tadpoles in rapidly drying ponds where *S. bombifrons* may fail to metamorphose (*1*, *2*). Moreover, tadpole size predicts patterns of adult size, and greater size can increase pre- and post-metamorphic survival (*10*–*12*). We based our study on these details, and predicted that *S. bombifrons* females would only express preferences for high-quality heterospecifics in environments and populations where hybridization confers a net advantage. Our predictions were met (*3*). Braun *et al.*'s criticisms cannot account for these behavior data nor their relationship to what is known about hybrid fitness in this system [(*3*) and references therein].

Braun *et al.* assert that our statistical analysis of fitness was inappropriate. They state that regression coefficients for fathers' pulse rate as a predictor of PC1 do not differ from zero in three out of five models when family 9 is removed. However, they fail to note that the two best-fit models are not only robust to removing family 9 [$P < 0.05$, ΔAICc = 2.70, compared to ≥3.54 for remaining models; table S3 of (*3*)], but also to removing families 4 and 15. Our results therefore did not depend on outlier families (see also Table 1).

Braun *et al.* additionally contend that PC1 is difficult to interpret. Setting aside that mass, SVL, and GS all loaded positively onto PC1 and explained 91.6% of the variation [table S1 of (*3*)], we examined the variables separately. These analyses support our original findings: Sire pulse rate is significantly inversely associated with size and development rate in the best-fitting, non-averaged models, which are robust to the removal of outlier families (Table 1). Given Braun *et al.*'s focus on condition, we examined condition measurements that are appropriate for amphibians, and found that condition is not associated with sire pulse rate (Table 1). However, as noted above, size and development rate are better assays of fitness—particularly survival—in the context of spadefoot hybridization (Fig. 1).

The natural history, known fitness consequences of hybridization, evidence that sexual signals predict hybrid fitness, and behavior data across the different environments and populations cannot be ignored. Taken together, they demonstrate adaptive mate choice among heterospecific males.

REFERENCES AND NOTES

1. K. S. Pfennig, Facultative mate choice drives adaptive hybridization. *Science* **318**, 965–967 (2007). doi:10.1126/science.1146035 Medline
2. K. S. Pfennig, M. A. Simovich, Differential selection to avoid hybridization in two toad species. *Evolution* **56**, 1840–1848 (2002). doi:10.1111/j.0014-3820.2002.tb00198.x Medline
3. C. Chen, K. S. Pfennig, Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates. *Science* **367**, 1377–1379 (2020). doi:10.1126/science.aaz5109 Medline
4. M. J. Braun, G. S. Wilkinson, B. S. Cade, Comment on "Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates". *Science* **370**, eabd3905 (2020).
5. Braun *et al.* state that we excluded three families with zero survivorship from our analyses [despite their note 4], but they confound survivorship with absence of offspring: One female did not produce eggs, and the other two produced few eggs, none of which developed. Including these families would conflate female fecundity with offspring survival. Notably, Braun *et al.* also exclude these families, so their rationale for including them in their figure 1A is unclear.
6. G. G. Rosenthal, *Mate Choice* (Princeton Univ. Press, 2017).
7. A. M. Achorn, G. G. Rosenthal, It's not about him: Mismeasuring 'good genes' in sexual selection. *Trends Ecol. Evol.* **35**, 206–219 (2020). doi:10.1016/j.tree.2019.11.007 Medline
8. When we analyze survival as family-level proportions and include parent measures using a linear model, we find that the top model contains only a quadratic term of pulse rate, the second-best contains only sires' condition (ΔAICc = 3.00), and the third-best was the null model (ΔAICc = 3.86).
9. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, ed. 2, 2002).
10. D. W. Pfennig, A. Mabry, D. Orange, Environmental causes of correlations between age and size at metamorphosis in *Scaphiopus multiplicatus*. *Ecology* **72**, 2240–2248 (1991). doi:10.2307/1941574
11. K. S. Pfennig, D. W. Pfennig, Character displacement as the "best of a bad situation": Fitness trade-offs resulting from selection to minimize resource and mate competition. *Evolution* **59**, 2200–2208 (2005). Medline
12. J. E. Earl, H. H. Whiteman, Are commonly used fitness predictors accurate? A meta-analysis of amphibian size and age at metamorphosis. *Copeia* **103**, 297–309 (2015). doi:10.1643/CH-14-128
13. K. Bartoń, R package version 1.42.41 (2018).
14. J. G. MacCracken, J. L. Stebbings, Test of a body condition index with amphibians. *J. Herpetol.* **46**, 346–350 (2012). doi:10.1670/10-292

ACKNOWLEDGMENTS

We thank J. Umbanhowar for statistical consultation, and D. Pfennig, J. Kingsolver, G. Calabrese, P. Kelly, E. Laub, M. Reynolds, and B. Loflin for discussion. Supported by NSF grant IOS-1555520; C.C. was also supported by the NSF GRFP (DGE-1650116).

10 July 2020; accepted 10 September 2020
Published online 9 October 2020
10.1126/science.abd5373

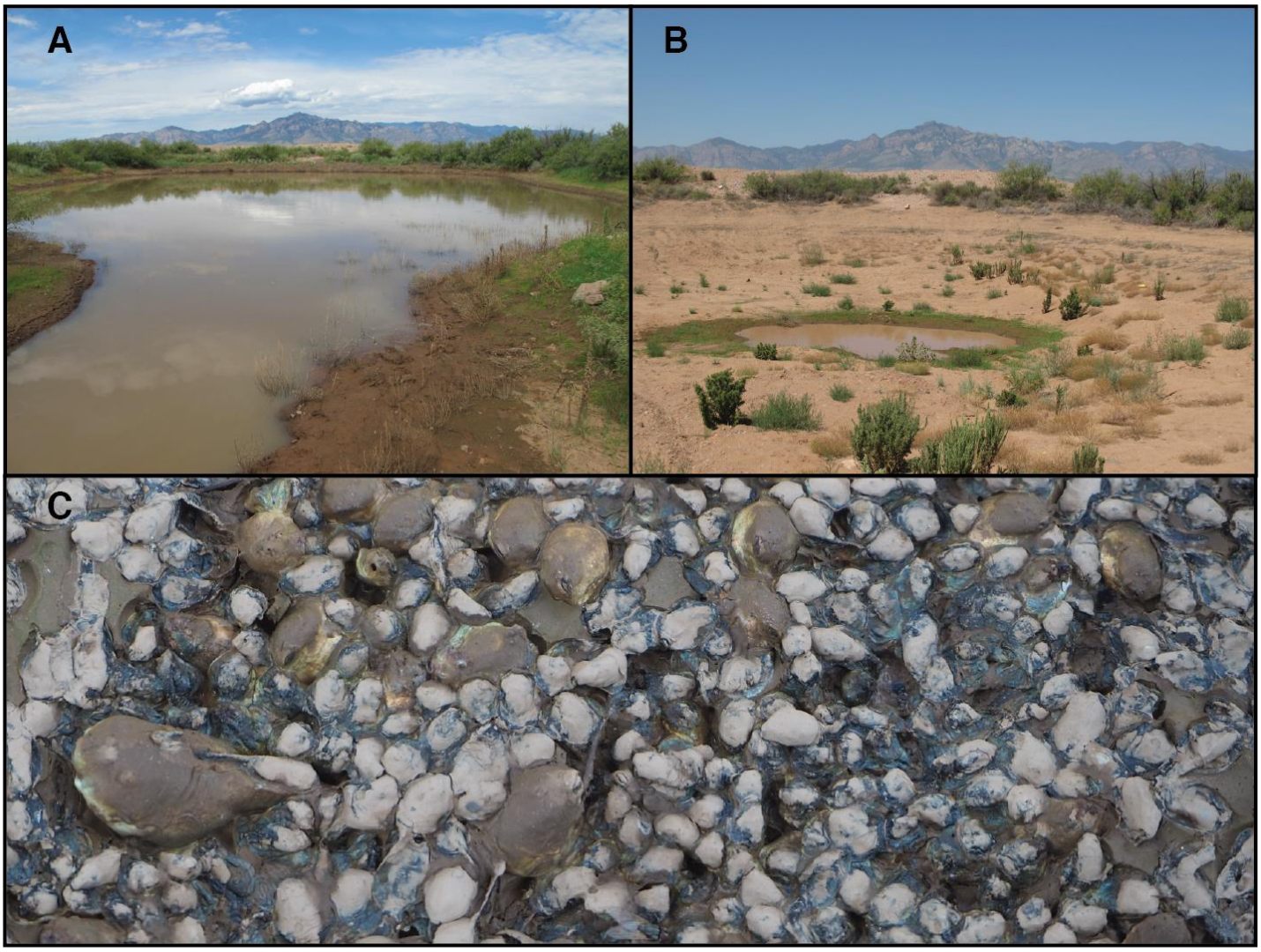


Fig. 1. Spadefoot breeding ponds and tadpoles. Ponds vary in depth and duration. (A) In wet years, when ponds last longer, slowly developing *S. bombifrons* tadpoles have time to metamorphose. (B) In dry years, the same pond can dry before *S. bombifrons* tadpoles metamorphose. (C) Tadpoles die if they fail to metamorphose (here, all "pebbles" are dead or dying tadpoles), but larger tadpoles desiccate more slowly and could survive if ponds refill from rain. Because hybrids develop faster than *S. bombifrons*, female *S. bombifrons* have higher net fitness by hybridizing in shallow ponds (*1*, *2*). [Photo credits: David Pfennig, (A) and (B); Karin Pfennig (C)]

Table 1. Summary of relationships between our individual fitness proxies and sire pulse rate. Non-nested candidate models were obtained using dredge from "MuMIn," and all contained family as a random effect (*13*). We refer to models with regression coefficients for sire pulse rate that were significantly different from zero ($P < 0.05$) as "top models," because all of those models had smaller ΔAICc scores than models that found no significant relationship with pulse rate or that did not include pulse rate. All of the top models identified inverse relationships with sire pulse rate. Other models did not contain sire pulse rate or found no significant relationship with sire pulse rate; the single exception identified a significant inverse relationship between GS and sire pulse rate (ΔAICc = 4.60), but we do not refer to it as a "top model" because two models that did not contain pulse rate had smaller ΔAICc values. Marginal R^2 of pulse rate for linear mixed-effect models was obtained using r.squaredGLMM in package "MuMIn" (*13*); we do not provide R^2 for GS because pseudo-R^2 values for ordered logit models are not comparable to those for linear regressions. We examined our top models without three families that had both low survival and sires with more extreme values of pulse rate: family 9 (sired by the male with the fastest pulse rate) and families 4 and 15 (sired by the males with the slowest pulse rates). We report the number of models that were robust to their removal [i.e., those that still found that regression coefficients for pulse rate were significantly different from zero ($P < 0.05$)]. SVL and mass are conflated, but we present both for completeness.

Fitness proxy	Model type	Number of candidate models	Number of top models	Combined weight of top models	ΔAICc range of top models	ΔAICc range of other models	Marginal R^2 range of pulse rate in the top models	Number of top models robust to removal of family 9	Number of top models robust to removal of families 9, 4, and 15
SVL	Linear mixed-effect model	9	6	0.982	0–5.97	8.13–9.01	0.058–0.091	4	3
Mass	Linear mixed-effect model	11	8	0.989	0–5.73	9.34–9.72	0.060–0.096	6	6
Gosner stage	Ordered logit model	10	5	0.837	0–2.37	4.14–5.49	—	3	3
Condition*	Linear mixed-effect model	4	0	—	—	0–3.01	—	—	—
Condition†	Linear mixed-effect model	6	0	—	—	0–2.35	—	—	—

*Condition calculated using the residuals of a cubic regression of mass on SVL (*3*). †Condition calculated using the scaled mass index (*14*).

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

DEVELOPMENTAL BIOLOGY

Intersexuality in female moles

Female moles are intersexual and develop masculinizing ovotestes, a distinctive trait among mammals. Real *et al*. investigated the origin of this trait by sequencing the Iberian mole genome and applying comparative strategies that integrate transcriptomic, epigenetic, and chromatin interaction data. They identified mole-specific genomic rearrangements that alter the three-dimensional regulatory landscape of the androgen-converting gene *CYP17A1* and the pro-testicular factor gene *FGF9*, both of which show distinct expression patterns in mole gonads. The use of transgenic mice confirms the capability of these factors to increase circulating testosterone levels and to induce gonadal masculinization. This study highlights how integrative approaches can reveal the phenotypic impact of genomic variation. —BAP *Science*, this issue p. 208

An Iberian mole peeks out of a burrow. Female moles exhibit intersex characteristics that are due to changes in chromosome structure and gene regulation.

VACCINES

Immunity after the flu shot

The seasonal flu shot is currently recommended each year because the influenza viral strains in circulation are continuously changing and because the antibody responses produced by the vaccine decline over time. In a human study of healthy volunteers, Davis *et al*. tracked antibody responses after flu vaccination. They investigated whether the vaccine led to the generation of antibody-secreting plasma cells in the bone marrow, a lymphoid organ that supports the survival of these cells for years. Although vaccination did generate influenza-specific cells, most were short-lived and lost within 1 year. The fact that a small number did persist over 1 year raises prospects that the longevity of flu vaccines can be improved and provides key information for the development of universal vaccines against influenza. —PNK

Science, this issue p. 237

NEUROSCIENCE

Representing the past and the future

Adaptive behavior requires the ability to analyze experience both prospectively and retrospectively. How can forward-ordered neural activity facilitate the storage or expression of reverse-ordered sequences? Wang *et al*. used multitetrode recordings from many individual hippocampal CA1 neurons in rats while simultaneously recording field potentials expressing theta oscillations. Spatial representation in the place cell network oscillated between forward and backward sweeps within each theta oscillation. Backward replay was associated with theta peaks, whereas forward replay was associated with theta troughs. Most cells fell into one category, but some corresponding to deep-layer neurons showed bimodular responses. Backward replay was driven by entorhinal inputs, whereas forward replay was evoked by CA3 inputs. These are important insights into the underlying basis of coding future and past experiences. —PRS

Science, this issue p. 247

QUANTUM GRAVITY

Spacetime, reconstructed

Theories of holographic duality feature a correspondence between a gravitational system and a strongly interacting conformal field theory (CFT) living on the system's boundary. Through this correspondence, the CFT encodes the geometry of spacetime in the gravitational system. Van Raamsdonk analyzed the role of entanglement in this theoretical framework. Instead of considering a single CFT, the author's starting point was a collection of CFT "bits" that are mutually entangled but do not interact with one another. The spacetime that these bits collectively encode was then shown to be arbitrarily close to the one encoded by the original CFT, suggesting that entanglement plays a crucial role in the emergence of spacetime. —JS

Science, this issue p. 198

PHOTO: TONY EVANS/TIMELAPSE LIBRARY LTD./GETTYIMAGES

Published by AAAS

BATTERIES

Engineering phosphorous anodes

A focus of battery research has been the development of a range of lithium, sodium, and potassium cathodes, but improving anode materials is also an important goal. Silicon has shown some promise for replacing graphite because of its exceptional capacity, but the dramatic volume change during lithiation-delithiation processes often leads to failure. Jin *et al*. developed a composite that is made of black phosphorous and graphite in its core and covered with swollen polyaniline. In contrast to previous efforts, bonding between the carbon and phosphorous allows for a high charging rate without sacrifices in capacity and cycling stability. —MSL

Science, this issue p. 192

ECOSYSTEM RECOVERY

Successful restoration of seagrass beds

Since 1999, broadcasting more than 70 million seeds of the seagrass *Zostera marina* into coastal lagoons in Virginia has led to the reestablishment of more than 3600 hectares of previously lost seagrass beds. Orth *et al*. found that seagrass recovery led to rapid increases in the production and diversity of animals, increases in carbon and nitrogen sequestration, decreases in turbidity, and increases in previously depleted bay scallops. This is one of the larger success stories of ecosystem restoration and demonstrates that reversal of ecological collapse is possible when sustained initial efforts lead to positive feedbacks that support ecosystem recovery. —MEH

Sci. Adv. 10.1126/sciadv.abc6434 (2020).

PLANT SCIENCE

Building a safe space for stem cells

The meristem, the collection of stem cells that builds plants, is resistant to viral infection. Wu *et al*. now show that WUSCHEL, a transcription factor that helps to sustain stem cell production in the shoot apical meristem of *Arabidopsis*, also protects that stem cell domain from viruses. WUSCHEL inhibited viral protein synthesis by repressing methyltransferases that regulate ribosomal RNA processing and ribosome stability. —PJH

Science, this issue p. 227

TINNITUS

Relief from ringing

There is no universally effective treatment for tinnitus, which is phantom perceived sound or ringing in the ears. Conlon *et al.* studied the effects of bimodal neuromodulation using a device that delivers electrical stimulation to the tongue and sound to the ears. Participants reported reduced symptom severity at the end of 12 weeks of treatment, which persisted for 12 months. Results support the safety and potential utility of bimodal neuromodulation for tinnitus. —CC

Sci. Transl. Med. **12**, eabb2830 (2020).

PHOTO (LEFT TO RIGHT): NOAA CC BY 2.0; RJ STYLES/ALAMY STOCK PHOTO

Eel grass restoration off the coast of eastern Virginia has enabled repopulation of important species such as the bay scallop, pictured here.

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

An intermediate state between clear and cloudy sky greatly affects the radiative balance of the planet.

CLOUD PROPERTIES

Between states

The greatest obstacle to understanding and projecting climate warming is uncertainty about how clouds affect Earth's radiative balance. Clouds can have either cooling or warming effects depending on type and location. However, the sky is not either cloudy or clear; there is also an intermediate state with its own radiative properties. Eytan *et al.* analyzed satellite data about this "cloud twilight zone" and found that it has a significant effect on longwave radiative forcing. Its impact is equivalent to that of a 75 parts per million increase in atmospheric carbon dioxide, which is nearly equivalent to that caused by anthropogenic emissions for the past 60 years. —HJS *Nat. Geosci.* **13**, 669 (2020).

NEURODEVELOPMENT

Speech processing and reading skills

People differ in their ability to parse language in noisy environments. Especially challenging are settings in which the talking face is not visible and background noise consists of other people talking. Destoky *et al.* investigated how variation in speech processing affects the ability to read for children with typical reading ability and those with dyslexia. A child introduced to a spoken word in a noisy classroom setting may find that word difficult to recognize in the written context. Reciprocally, reduced reading experience would give a child a more limited mental lexicon

Published by AAAS

SYMBIOSIS

Hyphal toll roads through the soil

Filamentous fungi such as *Aspergillus nidulans* grow by hyphal branching and extension in a multicellular network. These networks make "highways" through the soil for motile bacteria such as *Bacillus subtilis*, but whether this relationship is exploitative or mutualist is not well understood. Abeysinghe *et al.* found that when *B. subtilis* grows together with *A. nidulans*, the bacteria's thiamine biosynthesis operon is induced. However, the fungal equivalent is repressed, and the authors presume that this is to save metabolic costs. If the fungal operon is experimentally ablated, the resulting growth defect of the fungus can be rescued by intact *B. subtilis*. It appears that bacteria using the fungal highway to reach new foraging grounds pay for the ride by delivering thiamine to the hyphal tips. —CA *Life Sci. Alliance* **3**, e202000878 (2020).

Mycelium of cultured *Aspergillus nidulans*. In the soil, this fungus allows bacteria to travel along its hyphae in return for delivering vitamins to help it grow.

to draw upon when trying to interpret what the teacher says in that noisy classroom. The authors suggest that small alterations in a variety of processing abilities, as reflected by cortical representation of speech in noise, determine reading abilities and differentiate those with dyslexia from typical readers. —PJH

PLOS Biol. **18**, e3000840 (2020).

BIOMEDICINE

Eating the heart healthy

The heart is composed of cardiomyocytes that beat all day every day with very little cellular turnover. How do these cells manage to keep themselves healthy in the face of such constant stress? Nicolás-Ávila *et al.* found that macrophages residing in the myocardium actively contribute to cardiomyocyte maintenance. Cardiomyocytes expel damaged organelles, including mitochondria, in packages that resemble structures in the brain called exophers. These membranous particles are then consumed by macrophages using phagocytosis. When cardiac macrophages were depleted in mice, the myocardium became compromised. Cardiomyocytes accumulated defective mitochondria and displayed activation of inflammasomes, autophagy, and metabolic dysfunction. Thus, cardiac macrophages play an important role in maintaining heart health by providing a way for cardiomyocytes to maintain organelle and metabolic homeostasis under incessant stress. —SMH

Cell 10.1016/j.cell.2020.08.031 (2020).

NEUROSCIENCE

Getting a GRIP on AMPARs

Regulation of synaptic AMPA receptors (AMPARs) is a key process implicated in synaptic plasticity. Glutamate receptor interacting protein 1 (GRIP1) is a scaffolding protein that interacts directly with several members of the AMPAR family. GRIP1 regulates the surface expression and synaptic stabilization of AMPARs, but its role in the expression of long-term potentiation (LTP) is unknown. Tan *et al.* investigated the function of GRIP1 in LTP and its impact on learning and memory. GRIP1 was recruited into synapses with AMPARs. The GRIP1-AMPAR interaction was enhanced during LTP. Loss of GRIP1 blocked the activity-induced accumulation of synaptic AMPARs and mice deficient in Grip1 exhibited learning and memory deficits. These findings indicate an essential role for GRIP1 in synaptic plasticity and cognitive functions. —PRS

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.2014827117 (2020).

QUANTUM CHEMISTRY

UV-vis spectra from time-independent DFT

Theoretical chemistry methods used to simulate electronic absorption properties, such as time-dependent density functional theory (TD-DFT), are usually applicable only to small- or medium-size systems because of their unfavorable scaling behavior. Chan and Hirao propose an orbital energy-based approximation scheme aimed to accurately simulate ultraviolet-visible (UV-vis) spectra using a lower-cost conventional (time-independent) version of DFT. Their proposed approach demonstrates reliable performance across an extensive range of examples, from small organic molecules to large metal-organic frameworks with electrochromic properties. Its accuracy is comparable to TD-DFT but is achieved with just two single-point conventional DFT calculations. —YS

J. Phys. Chem. Lett. **11**, 7882 (2020).

ECONOMIC INEQUALITY

Minimum wage policy and the racial earnings gap

The earnings gap between white and Black workers in the United States fell substantially only once over the past 70 years. Derenoncourt and Montialoux show how this change was driven in large part by the 1967 extension of the federal minimum wage to sectors that employed about 20% of the total U.S. workforce but nearly a third of Black workers, including agriculture, hotels, and schools. This policy extension can explain more than 20% of the decline in the racial earnings gap between 1965 and 1980 and did not have large adverse employment effects on either Black or white workers. The gap between the average wage in the treated industries and the rest of the economy fell, as did the gap between Black and white workers within the treated industries. —BW

Q. J. Econ. 10.1093/qje/qjaa031 (2020).

PHOTO: GEOFF TOMPKINSON/SCIENCE SOURCE

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

BIOTECHNOLOGY

Evolution of bioengineered blood vessels

Biotechnology approaches to repair and replace arteries have been under development for more than a century. Early synthetic approaches used rubber-based replacements, which then evolved into the use of polymer fabrics and, more recently, into biological approaches that permit the growth of blood vessels in the laboratory. Niklason and Lawson review the scientific and technological advances that allow the regeneration of a patient's own blood vessels. The authors discuss how blood vessel cells, when combined with suitable substrates for tissue growth under conditions that mimic human physiology, can produce functional bioengineered arteries. These biological approaches pave the way to advancing how vascular disease is managed and treated in the future. —PNK

Science, this issue p. 185

ELECTROCHEMISTRY

Electrolysis feels the heat

Electricity infrastructure powered by sunlight and wind requires flexible storage capacity to compensate for the intermittency of these sources. In this context, Hauch *et al*. review progress in solid oxide electrolyzer technology to split water and/or carbon dioxide into chemical fuels. These devices, which rely on oxide conduction between cathode and anode, use nonprecious metals as catalysts and operate above 600°C, thereby benefiting from thermodynamic and kinetic efficiencies. The authors highlight recent optimizations of cell components as well as systems-level architecture. —JSY

Science, this issue p. 186

CORONAVIRUS

Questions about bats and viruses

Bats have been attributed as the source of several recent diseases that have made the jump from animal hosts to humans. It has been suggested that bats have an attenuated inflammatory immune response that allows them to tolerate more viruses than other mammals, making them a hotbed for zoonotic virus evolution. In a Perspective, Streicker and Gilbert discuss what is known about bats as a source of zoonotic viruses, raising questions about the influence of species diversity on such generalizations. Moreover, zoonotic spillover is a poorly understood and complicated process that is influenced by anthropogenic behaviors such as culling, the wildlife trade, and consumption. The role of bats in zoonoses needs much more investigation, including long-term monitoring, before solutions can be found. —GKA

Science, this issue p. 172

HIV

HIV-1 replication and integration in vitro

To infect a host cell, HIV-1 must reverse transcribe its single-stranded RNA genome into a double-stranded DNA copy and integrate that copy into a host chromosome. Reverse transcription and integration have been characterized separately but have not been reconstituted together outside of the cell. Christensen *et al*. now report that viral core particles can complete full reverse transcription and integration in a cell-free system. The external capsid shell of the core is required for efficient reverse transcription, and the replicating DNA can loop out of capsid openings. Integration requires cell extract, and this cell-free system should be useful for analyzing how host factors contribute to the first half of the HIV life cycle. —SMH

Science, this issue p. 187

VIRAL INFECTIONS

ACHOO!-trophils?

Why some individuals come down with a cold in any given year, whereas others are spared, is poorly understood. Habibi *et al*. exposed volunteers to respiratory syncytial virus (RSV), one of the pathogens responsible for the common cold, and then followed them over the ensuing 2 weeks (see the Perspective by Mirchandani and Walmsley). The main factor distinguishing the infected from the noninfected was that the former showed signs of airway neutrophil activation before exposure to RSV. Conversely, protected individuals showed enhanced interleukin-17 signaling soon after virus inoculation. Similarly, mice pretreated with a neutrophil chemoattractant were more susceptible to RSV infection and $CD8^+$ T cell–driven disease. Neutrophil proteases, which can modulate cytokine activity, may explain the disparity and could provide therapeutic targets for RSV and other respiratory infections. —STS

Science, this issue p. 188; see also p. 166

STRUCTURAL BIOLOGY

A moving target

Abl kinase is an important signaling protein that is dysregulated in leukemia and other cancers and is the target of inhibitors such as imatinib. Like other kinases, Abl kinase is dynamic, and regulating conformational dynamics is key to regulating activity. Xie *et al*. used nuclear magnetic resonance to show that the Abl kinase domain interconverts between one active and two inactive states. Imatinib stabilizes an inactive conformation, and several resistance mutations act by destabilizing this conformation. In a construct that includes the regulatory domain, depending on the relative arrangement of the kinase and regulatory domains, the kinase domain is stabilized in either the active state or one of the inhibited states. Understanding the conformational dynamics of kinases can be leveraged to design selective drugs. —VV

Science, this issue p. 189

METALLURGY

Disentangling fatigue

Metal fatigues when repeatedly loaded, ultimately failing when cracks form and propagate through the material. Lavenstein *et al*. studied the origins of this process in nickel. Using high-resolution observations, they tracked how dislocations evolved into microstructural features called persistent slip bands that preceded crack formation. The evolution of tangles of dislocations to a more regularly spaced pattern form the basis for the persistent slip bands and provide a road map for understanding fatigue cracking in metals. —BG

Science, this issue p. 190

NEUROSCIENCE

Interface targeting skirts excitotoxicity

Nearly all attempts to use traditional *N*-methyl-D-aspartate receptor (NMDAR) antagonists to treat neurodegenerative diseases have failed. This is because NMDARs are not only promoters of neuronal death but also have essential physiological roles in synaptic plasticity and cognitive functions such as learning and memory. Yan *et al*. explored the structural basis of NMDAR coupling to neuronal cell death (see the Perspective by Jones). The death-promoting activity, but not the essential physiological function, was mediated by the physical interaction of NMDARs with TRPM4, a calcium-impermeable ion channel activated by

Published by AAAS

intracellular calcium, depolarization, and temperature. A subsequent structure-based computational drug screen led to the discovery of neuroprotective small molecules that block the NMDAR-TRPM4 interaction interface but spare the critical healthy NMDAR function. —PRS

Science, this issue p. 191; see also p. 168

CORONAVIRUS

Flexible spikes

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein enables viral entry into host cells by binding to the angiotensin-converting enzyme 2 (ACE2) receptor and is a major target for neutralizing antibodies. About 20 to 40 spikes decorate the surface of virions. Turoňová *et al*. now show that the spike is flexibly connected to the viral surface by three hinges that are well protected by glycosylation sites. The flexibility imparted by these hinges may explain how multiple spikes act in concert to engage onto the flat surface of a host cell. —SMH

Science, this issue p. 203

SUPERCONDUCTIVITY

Alternative route to a 2D superconductor

Single layers of transition metal dichalcogenides exhibit exotic properties, including superconductivity. The usual route to obtaining such samples is to exfoliate a three-dimensional (3D) crystal. Devarakonda *et al*. instead grew a superlattice comprising alternating layers of the transition metal dichalcogenide hexagonal NbS_2 and the material Ba_3NbS_5 (see the Perspective by Schoop). The inert Ba_3NbS_5 layers serve to dissociate the superconducting NbS_2 layers from one another, resulting in 2D superconductivity with high carrier mobility. The combination of high mobility and reduced dimensionality may give rise to exotic quantum phases. —JS

Science, this issue p. 231 see also p. 170

ORGANIC CHEMISTRY

A potential method to tune reactivity

In organic chemistry, reaction rates are often sensitive to substituents that either donate or withdraw electron density from a reactive carbon center. Heo *et al*. now report that when reactants are tethered to an electrode, the polarization effect of an applied potential can exert an analogous influence, essentially acting as a tunable functional group. The authors demonstrate rate modulations in ester hydrolysis and arene cross-coupling, as well as a two-step acid-to-amide conversion that benefits from inversion of the polarization midway through the reaction. —JSY

Science, this issue p. 214

NOVEL COMMUNITIES

Change begets change

In the Anthropocene, humans are altering ecosystems, causing extinctions, and reassorting species distributions. As we facilitate these changes, we are creating new collections of species. Such "novel communities" are not specific to our epoch, and the patterns of diversity and extinction associated with past events can shed light on the implications of current community restructuring. Pandolfi *et al*. looked at marine plankton communities over the past ~66 million years and found that the emergence of novel communities leads to further novelty and extinction (see the Perspective by Dornelas and Madin). Although community change is a natural biological response to environmental change, the current rate of change could lead to impending and rapid impacts. —SNV

Science, this issue p. 220; see also p. 164

CORONAVIRUS

The CIITAdel keeps viruses at bay

A better understanding of cellular mechanisms involved in viral resistance is needed for the next generation of antiviral therapies. Bruchez *et al*. used a transposon-mediated gene-activation screen to search for previously unreported host restriction factors for Ebola virus (see the Perspective by Wells and Coyne). The authors found that a transcription factor, major histocompatibility complex class II transactivator (CIITA), induces resistance in human cell lines by directing the expression of the p41 isoform of the invariant chain (CD74). CD74 p41 then disrupts cathepsin-mediated Ebola glycoprotein processing, which prevents viral fusion and entry. CD74 p41 can also stymie the endosomal entry of coronaviruses, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This work should inform future treatments against cathepsin-dependent viruses such as filoviruses and coronaviruses. Additionally, the screening strategy used may serve as a blueprint for uncovering resistance mechanisms against other dangerous pathogens. —STS

Science, this issue p. 241 see also p. 167

ELECTRON MICROSCOPY

Going for the gold

Single-particle cryogenic electron microscopy (cryo-EM) has become a go-to technique for structural biologists. Although data-processing and reconstruction methods have improved, innovations in sample preparation and data collection are essential to reliably achieve high-resolution reconstructions while also reducing the amount of time required per structure. Naydenova *et al*. tackled the issue of electron beam–induced particle movement, a major source of information loss, by designing a gold sample support that prevents buckling of the extremely thin layer of ice in which the particles are suspended (see Perspective by Rapp and Carragher). The negligible particle displacement permits extrapolation to "zero exposure" structure factors, revealing features typically lost in cryo-EM structures. Far fewer particles per unit resolution are required, which greatly accelerates structure determination, especially at high resolution. —MAF

Science, this issue p. 223; see also p. 171

CANCER IMMUNOLOGY

TAMs reprogram tumor metabolism

Macrophages in the tumor microenvironment (called tumor-associated macrophages or TAMs) often support tumor growth and metastatic progression. Gomez *et al*. found that secretion of the cytokine transforming growth factor–beta (TGF-β) from anti-inflammatory TAMs suppressed the abundance of succinate dehydrogenase in breast cancer cells. The loss of this metabolic enzyme promoted glycolysis, thereby enhancing tumor growth, angiogenesis, and immunosuppression. Depleting anti-inflammatory TAMs or blocking TGF-β suppressed these effects in mice. —LKF

Sci. Signal. **13**, eaax4585 (2020).

VIRAL IMMUNOLOGY

Revenge against rotavirus

Rotavirus (RV) is an enteric infection that can cause severe morbidity and contributes to childhood mortality in resource-poor countries where the existing vaccine is of moderate efficacy. A previous study showed that the cytokines interleukin-18 (IL-18) and IL-22 work together to prevent or treat RV infection. Zhang *et al.* now show that IL-22 drives intestinal epithelial cell (IEC) proliferation and migration toward villus tips, resulting in extrusion of differentiated IECs, which are sites of RV replication. IL-18 induces cell death of infected IECs, and together these mechanisms accelerate clearance of RV from the intestine. These findings suggest that these cytokines may support the clearance of other enteric viral infections. —CNF

Sci. Immunol. **5**, eabd2876 (2020).

Published by AAAS

REVIEW SUMMARY

BIOTECHNOLOGY

Bioengineered human blood vessels

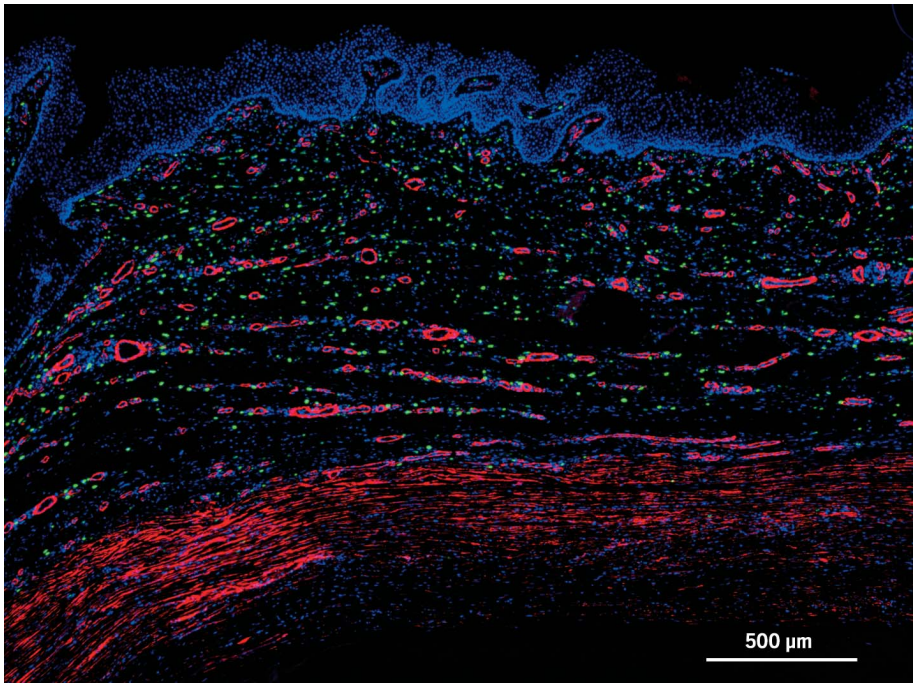
Laura E. Niklason* and Jeffrey H. Lawson*

BACKGROUND: Vascular replacement and repair for the treatment of atherosclerotic disease, infection, and traumatic injury are some of the most commonly performed surgical procedures in the Western world. In the United States alone, hundreds of thousands of coronary and peripheral arteries are repaired, replaced, or bypassed every year. But despite the enormity of the clinical need for engineered arterial replacements, the equally enormous simultaneous challenges of immune acceptance, requisite tissue mechanics, low thrombogenicity, and immediate availability have made the broad clinical application of engineered arteries quite difficult to achieve. In this regard, recent years have seen the fusion of cell biology, physiology, and engineering to now allow for the creation of human tissues that can truly function in the setting of vascular repair and replacement.

ADVANCES: For a biological engineered artery to function successfully without requiring immunosuppression, the following objectives should be met: (i) The engineered artery should have an extracellular matrix of sufficient quality to provide suitable tensile, suture retention, and rupture strength properties. A focus on production of suitable amounts of high-quality cross-linked vascular collagens type I and III is probably necessary for any biological engineered artery to be successful. (ii) To minimize risks of inflammation, foreign-body response, and immune recognition, the vascular tissue matrix should be of human origin and without substantial synthetic material additives or artificial covalent cross-linking. (iii) If the engineered artery is cellular, even if the cells are nonviable, those cells should be autologous to prevent immune recognition, degradation, and aneurysm formation in the implanted vessels. (iv) Once implanted, the engineered arteries should have the potential to be remodeled, repopulated, and rejuvenated by the host. (v) For small-caliber or low-flow arterial bypass applications, it is likely that a suitable nonthrombogenic luminal surface is required. This surface may be either cellular or biochemical, but it should prevent blood coagulation contact activation, platelet adhesion and activation, and thrombosis in the arterial system.

OUTLOOK: Guided by the design criteria above, engineered blood vessels have been developed by several groups that have progressed to clinical trials. Recent clinical studies have demonstrated the feasibility of using human tissue–engineered blood vessels in the settings of vascular trauma, peripheral arterial disease, and vascular access for hemodialysis. Engineered arteries reaching the clinical domain have been composed of autologous cells or allogeneic cells, or have been engineered from allogeneic cells or tissues and then decellularized. Vascular functionality in patients has been demonstrated in both low-pressure environments (pediatric cardiac surgery) and high-pressure environments (peripheral arterial surgery in adults).

Autologous cell approaches have shown some promise, particularly in clinical settings of venous reconstruction and low pressure and in pediatric populations. However, scaling production of engineered arteries to tens of thousands of vessels per year, as would be needed to treat arterial atherosclerosis at large scale, presents enormous logistical challenges if autologous cell sources are used. Hence, it is likely that future successes of engineered arteries will employ allogeneic human cells or cell banks to generate tissues at clinically relevant scales, and suitable strategies will be required to prevent adaptive immunity and rejection of these vessels. Furthermore, next-generation techniques such as three-dimensional bioprinting of both cells and matrix may one day allow vessel production at accelerated speeds, possibly producing usable tissues in hours or days, rather than weeks or months. Microvascular and cardiac tissue engineering are also making important strides, pointing toward a future that could enable revascularization of solid organs. The evolution of scientific thinking and approaches that have brought us to this point is summarized in this review. ■



An engineered human artery cultured from human vascular cells and implanted into a patient. Immunostaining for smooth muscle (red), progenitor cells (green), and cell nuclei (blue) shows extensive cellular repopulation of the engineered vessel. Engineered cells were implanted into the patient for 4 years. The layer of red-staining cells at the bottom of the image shows the repopulated engineered vessel wall. Blue staining at the top shows the nuclei of skin cells.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: laura.niklason@yale.edu (L.E.N.); lawson@humacyte.com (J.H.L.)
Cite this article as L. E. Niklason, J. H. Lawson, *Science* **370**, eaaw8682 (2020). DOI: 10.1126/science.aaw8682

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aaw8682

REVIEW

BIOTECHNOLOGY

Bioengineered human blood vessels

Laura E. Niklason[1,2]* and Jeffrey H. Lawson[2,3]*

Since the advent of the vascular anastomosis by Alexis Carrel in the early 20th century, the repair and replacement of blood vessels have been key to treating acute injuries, as well as chronic atherosclerotic disease. Arteries serve diverse mechanical and biological functions, such as conducting blood to tissues, interacting with the coagulation system, and modulating resistance to blood flow. Early approaches for arterial replacement used artificial materials, which were supplanted by polymer fabrics in recent decades. With recent advances in the engineering of connective tissues, including arteries, we are on the cusp of seeing engineered human arteries become mainstays of surgical therapy for vascular disease. Progress in our understanding of physiology, cell biology, and biomanufacturing over the past several decades has made these advances possible.

The field of regenerative medicine has been growing steadily since the concept of tissue engineering was first proposed in the 1980s (*1*). The development of cell-based therapies (beyond bone marrow transplantation) has been growing progressively in recent decades, with more than 600 cell-based or engineered tissue–based clinical trials currently ongoing (www.clinicaltrials.gov). Engineered tissue replacements can be parsed into connective tissues, which rely primarily on the extracellular matrix for mechanical properties, and solid organs, which rely on their constituent cells to perform critical physiological functions (*2*). Connective tissues, including cartilage, bone, tendon, ligament, skin, heart valve, and blood vessels, have been the subject of long-standing and comprehensive tissue-engineering efforts. Indeed, current U.S. Food and Drug Administration–approved and marketed products include engineered cartilage and skin (e.g., Dermagraft, Apligraf, and MACI), and clinical trials are active in most of the other connective tissue types listed.

Of these connective tissues, engineered blood vessels have received close attention. Engineered arteries require particularly exacting design criteria, including adequate tensile and recoil strengths to withstand billions of cardiac cycles while maintaining a resistance to thrombogenicity, immunogenicity, inflammation, and foreign-body response that can lead to conduit failure. Native arteries are composed of three layers: (i) an intima containing endothelium, sparse myofibroblast-like cells, and an internal elastic lamina; (ii) a medium containing vascular smooth muscle cells (SMCs) and concentric layers of collagens type I and III, as well as elastin lamellae; and (iii) an adventitia that contains loose connective tissue fibroblasts, microvasculature, and afferent neurons. Very broadly, functional endothelium in native vessels produces substances such as nitric oxide, thrombomodulin, and tissue plasminogen activator to prevent thrombus formation or to degrade luminal thrombus once it forms. Mechanical integrity and tensile properties of vessels are conferred by the fibrillar collagens and elastin, with the modulus of collagen being approximately 100 times higher than that of elastin. Lastly, contractility of blood vessels and their ability to modulate diameter in response to exogenous soluble factors, neural inputs, and flow are mediated by the smooth muscle layer.

When implanted into the body, engineered arteries must be immediately functional, without needing any "maturation period" to become fully competent in terms of mechanical strength or inflammatory properties. Acute thrombosis of an engineered artery can lead to tissue ischemia and the risk of loss of life or limb. Mechanical failure of an engineered artery, manifesting as tearing at the suture line or an acute rupture, can also be limb- or life-threatening for the patient because of anastomotic disruption, false aneurysm formation, and bleeding. Furthermore, different anatomic locations for engineered vessels have different design requirements: small-caliber and low-flow vessels are more subject to thrombosis than are large-diameter vessels, which in contrast require greater wall tensile properties for mechanical durability. Lastly, the growth potential of engineered vascular replacements is important, given the increasing use of reconstructive surgery for infants and children with cardiovascular anomalies (*3*).

This review provides a chronological summary of the technologies used for vascular engineering and replacement, with a focus on biological—as opposed to synthetic—vascular conduits. Accumulated learnings in the fields of vascular cell biology, blood coagulation biochemistry, cardiovascular surgery, and transplantation medicine have combined to enable us to culture engineered arteries that appear to be clinically functional and, in the long term, may change the practice of medicine.

Early work in vascular engineering

Alexis Carrel was a French surgeon practicing in the early 20th century, who was a pioneer in the fields of surgery, transplantation, and tissue and organ culture. Carrel received the Nobel Prize in Physiology or Medicine in 1912 for his development of a method to create a vascular anastomosis of native blood vessels, which opened the door to modern cardiovascular surgery and the possibility of solid organ transplantation (*4*) (Fig. 1A). Later, Carrel partnered with aviator and engineer Charles Lindbergh in the development of perfusion pumps and incubators to maintain organs in culture for prolonged time periods. The two men wrote a book on this topic in 1938, *The Culture of Organs*, wherein they presciently anticipated many of the apparatuses and approaches that are used in modern tissue engineering (*5*): "The incubator used should have an inner glass door to permit observation of the interior without cooling the apparatus…It is important to minimize the total length of rubber tubing on the perfusion pump side of the pulsation valve…The pulsation pressures are measured by a sphygmomanometer."

In the 1960s, Charles Sparks reported on several types of arterial replacements (*6*). At that time, artificial arteries made of silicone rubber were developed that could be attached to native vasculature by means of clamped connectors (Fig. 1B). However, Sparks rapidly pivoted to developing autologous "vessels," generated by implanting a tubular silicone rubber mandril into the leg of the intended patient and weeks later harvesting the resulting tubular fibrotic tissue tube to be used as an arterial replacement (*7*). Clinical trials of these "Sparks mandril" grafts continued into the 1970s but showed that the scar tissue–based tissue tubes tended to thrombose, or become aneurysmal over time (*8*, *9*). Use of the mandril grafts was abandoned in the 1980s, though a resurgence of this approach has recently been taken up by other investigators. For example, Rotmans and colleagues have used revised approaches wherein mandrils are pretreated with transforming growth factor–β (TGFβ), collagen, or plasma etching, in efforts to improve the foreign-body response to the implants, with the future intent to grow conduits for patients needing vascular access for hemodialysis (*10*).

[1]Departments of Anesthesiology and Biomedical Engineering, Yale University, New Haven, CT, USA. [2]Humacyte Inc., Durham, NC 27713, USA. [3]Department of Surgery, Duke University, Durham, NC, USA.
*Corresponding author. Email: laura.niklason@yale.edu (L.E.N.); lawson@humacyte.com (J.H.L.)

Concerted efforts at in vitro arterial engineering arguably began in the 1980s at the Massachusetts Institute of Technology (MIT), where Eugene Bell and colleagues reported the first culture of an engineered vessel made from vascular cells (*11*). Bovine vascular smooth muscle, endothelial, and fibroblast cells were suspended in a tubular collagen gel, which was remodeled over time in culture to form a cellular, tubular tissue. Rupture strengths of these conduits were only 90 mmHg at maximum, and the conduits required support by external Dacron sleeves to enable handling and to increase burst strength (Fig. 1C). Nonetheless, this team was the first to show the feasibility of culturing a macroscopic vascular analog in vitro.

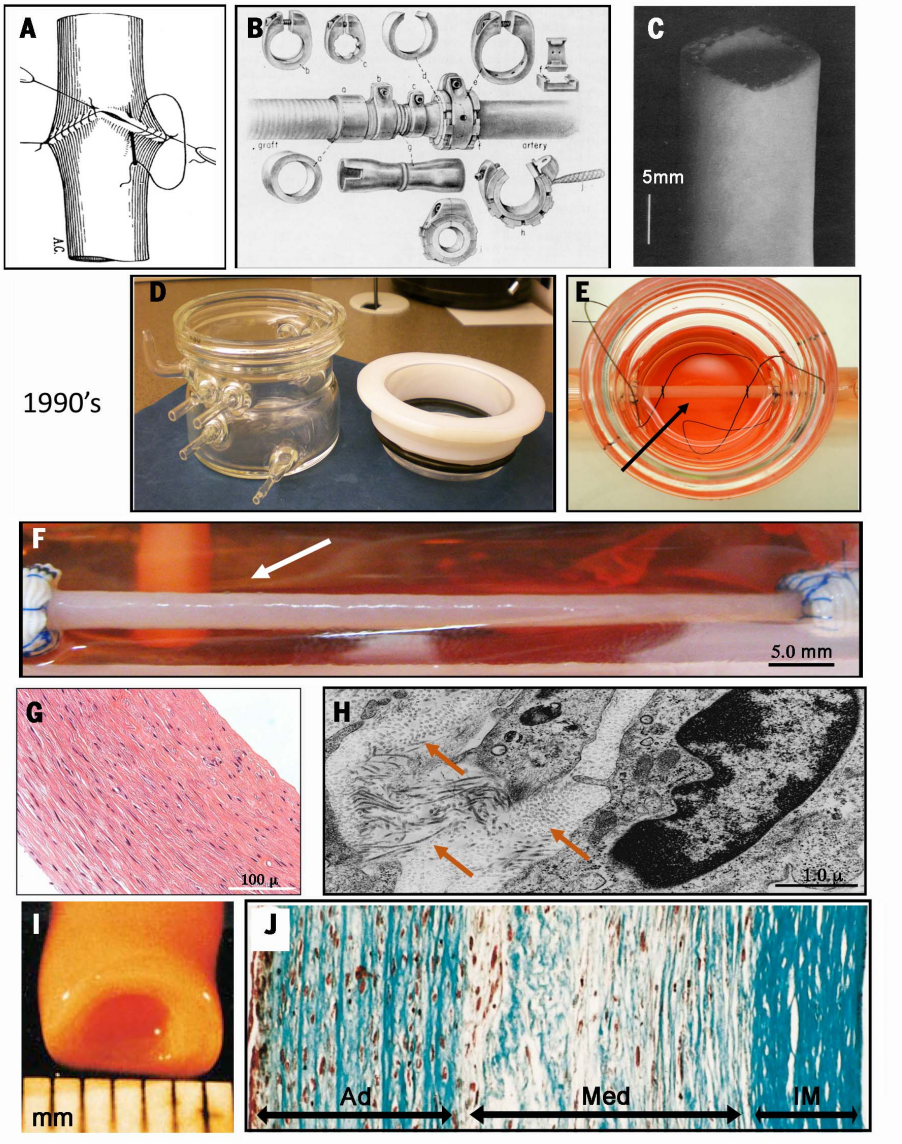


Fig. 1. Progression of engineered vascular replacements. (**A**) Drawing of the vascular anastomosis technique by Alexis Carrel (*73*). (**B**) Diagram of silicone-based arterial replacement reported by Charles Sparks in 1966 (*6*). (**C**) Gross photo of the original blood vessel model reported by Weinberg and Bell in 1998 [modified from (*11*)]. (**D** and **E**) Glass bioreactor with Teflon and rubber gasket gap, containing engineered bovine blood vessel 3 mm in diameter and 5 cm in length (black arrow). (**F**) Gross photo of 3-mm-diameter engineered blood vessel after 2 weeks of culture (white arrow). (**G** and **H**) Hematoxylin & eosin (H&E) cross-sectional image (G), and transmission electron micrograph image (H), of engineered bovine vessels showing ordered cellular layers (G) and synthesized collagen fibers [orange arrows in (H)]. (**I** and **J**) Gross photo (I) and Masson's trichrome stain (J) of engineered human blood vessel made from wrapped layers of SMCs and fibroblasts [modified from (*15*)]. In (J), adventitia ("Ad") is wrapped around media ("Med") and outside of the inner membrane ("IM," width ~125 μm).

Vascular engineering takes shape

In the wake of Bell's pioneering work proving the principle of arterial tissue regeneration in vitro, a large number of investigators took up the problem using a variety of approaches. In efforts to improve hemocompatibility of synthetic arterial grafts, Zilla, Deutsch, and colleagues developed means of seeding autologous endothelium into the lumen of expanded polytetrafluoroethylene (ePTFE) conduits (*12*–*14*). Endothelial cells (ECs) derived from small vein biopsies were expanded and cultured on ePTFE grafts over a period of roughly 4 weeks and then were implanted as infrainguinal femoropopliteal grafts in patients with severe peripheral arterial disease. In a study of more than 300 patients who received the endothelialized grafts, primary and secondary patency rates were comparable to those of native saphenous vein, making this endothelialization approach an important success (*14*). However, the need for harvesting suitable numbers of autologous ECs from vein segments, and challenges with long-term EC adherence to artificial surfaces, have likely prevented more widespread adoption of this technology.

The development of a completely biological, implantable vessel was first reported by L'Heureux and colleagues in 1998 (*15*). Fibroblasts and vascular SMCs were grown from human explant sources, cultured as sheets in the presence of ascorbate to support collagen deposition, and then wrapped sequentially around a 3-mm mandril to produce a tubular tissue that was endothelialized on its lumen. An "inner membrane" consisting of lyophilized fibroblast sheets provided enhanced strength to the tissue (Fig. 1, I and J). Over a culture period of 12 weeks, engineered arteries with rupture strengths >2000 mmHg were produced, having organized lamellae of cells and collagenous extracellular matrix, comparable to those of native arteries. The cultured human vessels maintained 50% patency for 1 week in a canine model of femoral artery grafting, laying the groundwork for eventual translation of these engineered arteries into the clinic (see "Vascular access for hemodialysis" below).

Contemporaneously, in the 1990s, a new paradigm was developed for arterial engineering that involved the culture of vascular SMCs and ECs, supported by rapidly degradable polymer scaffolds that were suspended within bioreactors providing pulsatile strain (Fig. 1, D to F) (*16*). The choice of rapidly degrading polyglycolic acid (PGA) scaffolds was deliberate, given the design goal of producing an implantable engineered tissue that would have little or no synthetic material that could serve as a nidus for inflammation or foreign-body response. Early work also showed the importance of physiological strain for the development of tissue mechanics and collagen matrix

deposition (17, 18), which was consistent with work of other investigators showing the impact of strain on SMCs (19, 20). Furthermore, careful supplementation of culture medium to support synthesis and cross-linking of extracellular matrix was pivotal for producing vessels with native-like rupture and suture retention strengths over the 8-week culture period. Histological and microscopic evaluation of engineered vessels showed dense lamellae of spindle-shaped SMCs, interspersed with organized layers of collagen fibrils but lacking elastin (Fig. 1, G and H). In 1999, the implantation of the first completely autologous engineered arteries into porcine recipients was reported. These vessels, which were 3 mm in diameter and 5 to 6 cm in length, showed patency and good mechanical durability for 4 weeks (18, 21). Derivatives of this technology have been subsequently translated into patients in the settings of hemodialysis access, peripheral arterial disease, and vascular trauma (discussed below).

The role of elastin in vascular mechanics and stability is an important one. More than 20 years ago, Li and colleagues demonstrated that mice lacking one or both copies of the tropoelastin gene developed an increase in lamellae of the aortic media, implying an important role for elastin in inducing quiescence of SMCs (22). Laboratory studies have shown that biaxial strain applied to SMCs grown on PGA mesh can result in some elastin deposition after 8 to 12 weeks of culture (23). In addition, several investigators have noted that fibrin may support more elastin deposition than do other scaffolds (24, 25). However, vascular grafts that are fashioned exclusively from elastin, or from elastin-mimetic biopolymers, tend to have low tensile strengths, making application in the arterial circulation a challenge thus far (26).

Vascular replacements for the venous system

Several strategies have been used to generate tissue-engineered conduits for the venous circulation. Although the repair and replacement of small-caliber veins are typically not clinically indicated, reconstruction of larger-caliber vessels such as the vena cava, iliac, and jugular veins can be beneficial for certain patients. Andreadis and colleagues developed engineered venous replacements using bone marrow–derived progenitor cells that were differentiated to express smooth muscle markers (25). When these cells were suspended in a tubular fibrin gel, the resulting conduits maintained mechanical integrity and developed evidence of elastin formation within weeks of implantation into the jugular veins of ovine recipients.

In humans, approaches have been developed by Shinoka and Breuer for the growth of larger-diameter conduits (10 to 20 mm) that have been used for the reconstruction of the low-pressure pulmonary outflow tract of infants born with hypoplastic left heart syndrome (27, 28). Using fibrous scaffolds composed of PGA, polylactic acid (PLA), polycaprolactone (PCL), and copolymers of PLA and PCL, Shinoka and Breuer have seeded autologous, bone marrow–derived mononuclear cells to produce an implantable conduit in the operating room. These conduits, when implanted into children requiring reconstruction of the central venous or pulmonary arterial systems, have generally functioned well. In 25 pediatric patients in Japan undergoing extracardiac total cavopulmonary connection using engineered vessels, all grafts were patent at 30 days (28). After 10 years of follow-up, roughly 30% of patients displayed some degree of graft stenosis, which was successfully treated with balloon angioplasty (29). No patients suffered aneurysm formation or graft rupture of the engineered vessels in this low-pressure implantation site. These results from Japan formed the basis for an ongoing U.S. phase 1 trial of similar conduits in patients with single-ventricle physiology, which was recently completed (30).

Other biological vascular replacements

Although creating biocompatible blood vessels from cells and tissue engineering has long been a goal, numerous approaches using both xenograft and allograft tissues have been studied. Xenografts, from both bovine and ovine sources, have been developed and used clinically (31). However, to avoid acute and/or chronic rejection of the xenogeneic conduit, the vascular material is chemically "fixed," making the tissue both immunologically and biologically inert (32, 33). Chemically cross-linked vascular matrix is resistant to remodeling by repopulating cells, and therefore cellular population of these conduits is very limited. Because the conduit cannot become "living" again, there is often progressive dilatation of the vascular tissue owing to protein degradation over time. Alternatively, calcification can be observed that is associated with the presence of fixed vascular SMCs, accompanied by local calcium precipitation.

Arterial grafts based on acellular xenogeneic matrix, such as porcine small intestinal submucosa (SIS), have shown good durability in sheep models both in vivo and ex vivo (34, 35). However, the use of SIS as a venous or arterial graft material in patients has been complicated both by stenosis and by acute mechanical disruption and pseudoaneurysm formation (36).

Cryopreserved, allogeneic, cadaveric human vessels have also been developed and used clinically. Although these vessels can be used in the acute setting, they are chronically susceptible to immune recognition, rejection, and aneurysmal degradation owing to the innate immunity in all of us (37, 38). Indeed, the survival of cardiac and kidney allografts is often limited not by rejection of the organ parenchyma but by adaptive immunity directed against the donor blood vessels (39). Allograft vasculopathy is thought to be a form of T helper cell 1–mediated delayed type hypersensitivity reaction, possibly to major histocompatibility complex class I and class II antigens on the surfaces of donor endothelial cells. In transplanted arteries, smooth muscle cell hyperplasia due to transplant vasculopathy occurs and can lead to luminal narrowing and occlusion, which may be due in part to interferon-γ derived from activated T lymphocytes (40). Although the level of expression of human leukocyte antigens in human smooth muscle cells is lower than that of endothelium, both cell types express antigens that can trigger adaptive immunity, even in cases where the cells have been killed but have not been removed from the vessel wall.

Fundamental design criteria for engineered arteries

Key to approaching the design of engineered vascular replacements is a focus on critical functional attributes and an avoidance of known (or predictable) failure mechanisms. Identifying the required arterial characteristics and functions allows tissue engineers to focus their energies on achieving these attributes while not being distracted by other, less critical features. Notably, this approach of identifying key functions differs in important ways from asserting that any engineered tissue must simply mimic all cellular, compositional, and physiological aspects of the corresponding native tissue. Indeed, the history of successful medical devices such as cardiac pacemakers, hip replacements, and cochlear implants teaches us that for "spare parts" to be effective, they need not replicate every single aspect of the original tissue.

For a biological engineered artery to function successfully without requiring immunosuppression, the following elements appear to be essential:

1) The engineered artery must have extracellular matrix of sufficient quality to provide suitable tensile, suture retention, and rupture strength properties. Because fibrillar cross-linked collagen has tensile properties that are roughly a hundred times greater than those of cells and extracellular elastin (2, 41), a focus on production of suitable amounts and quality of cross-linked vascular collagens type I and III should enable success for biological engineered arteries.

2) To minimize risks of inflammation, foreign-body response, and immune recognition, the extracellular matrix should be of human origin and without substantial synthetic material additives or artificial covalent cross-linking (42).

3) If the engineered artery is cellular, even if those cells are nonviable, the cells should be autologous to prevent immune recognition, matrix degradation, and aneurysm formation of the implanted vessels. If allogeneic or xenogeneic cells are used, then their complete antigenic removal is necessary before implantation, to avoid triggering immune recognition in the recipient (*39*).

4) Once implanted, the engineered arteries should have the potential to be remodeled by the host. Because the half-life of vascular collagens is on the order of months, population of engineered vessels with autologous cells that can maintain matrix integrity is needed to prevent long-term mechanical failure arising from collagen breakdown (*43*).

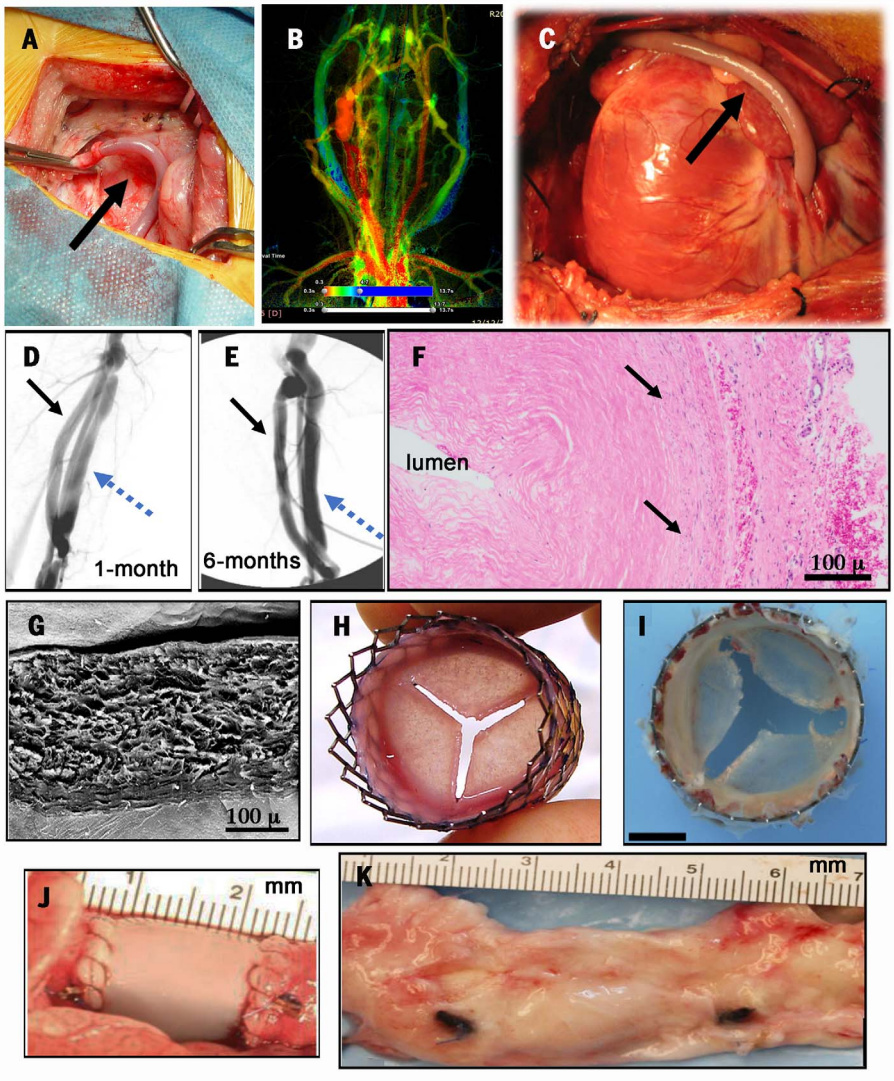


Fig. 2. Decellularized, engineered cardiovascular tissues. (**A** and **B**) Surgical photo (A) and perfusion scan (B) of an engineered, decellularized bovine vessel, 6 mm in diameter, inserted as an arteriovenous graft in a porcine recipient [arrow in (A)], between the carotid artery and jugular vein. (**C**) Engineered canine artery (arrow), decellularized but coated with autologous endothelium, implanted as a coronary bypass graft in a canine heart. (**D** to **F**) Engineered human vessels, 6 mm in diameter, implanted into baboon recipients as arteriovenous conduits (solid arrows). Panels (D) and (E) contain angiograms at 1 and 6 months after implantation (dotted arrows are dilated native vein downstream from arteriovenous conduit). Panel (F) shows an H&E image of an engineered human vessel explanted from a baboon after 6 months, with cell infiltration in the outer surface of the vessel (black arrows). (**G**) Scanning electron micrograph of decellularized engineered porcine artery, showing voids in the wall where smooth muscle cells once resided. (**H** and **I**) Engineered human decellularized heart valves, preoperatively (H) and after 8 weeks in the pulmonic valve position in a baboon [modified from (*51*)]. (**J** and **K**) Engineered, decellularized ovine vessel implanted into the pulmonary artery position in a lamb at age 8 weeks: at time of operation (J) and after 42 weeks of implantation, showing growth between anastomotic sites (*52*).

5) For small-caliber or low-flow arterial grafting applications, it is possible that a suitable luminal "blood contacting" surface will be required to reduce or prevent blood coagulation contact activation and platelet adhesion and activation, and to maintain vessel patency. This luminal surface may comprise endothelium derived from autologous vein or from circulating progenitors, or it may comprise other surface treatments that reduce thrombogenicity (*34*, *44*).

Acellular engineered cardiovascular conduits: Preclinical applications

Driven by the challenges inherent in autologous cell-based connective tissue engineering and armed with the insight that connective tissue mechanics is largely derived from highly conserved extracellular matrix proteins, research has explored whether engineered arteries could be decellularized. The choice of this strategy was driven by several of the key design criteria listed above and has the potential to produce engineered arteries that are available "off-the-shelf" and do not require a biopsy of cells from the patient. The strategy of decellularizing native connective tissues, such as heart valves and saphenous veins, has been in use for several decades and is utilized to produce clinically available valve homografts and vascular conduits, as well as skin grafts (e.g., Synergraft, Alloderm) (*45*–*47*). In 2003, the first report was published of a porcine engineered artery that had been grown from cells in vitro and then decellularized, showing that the mechanical properties of the artery were preserved after cell removal (*48*). This result opened the door to producing engineered blood vessels by using human allogeneic donor cells, which could then be chemically treated to remove the donor cells and antigens while leaving the mechanically robust extracellular matrix intact.

To test this concept in large animal models, acellular, engineered vessels grown from allogeneic donor cells were implanted into porcine recipients as arteriovenous conduits (Fig. 2, A and B) and into canine recipients as carotid and coronary artery bypass grafts (lumen seeded with autologous ECs, Fig. 2C) (*49*). Engineered arteries were grown from allogeneic vascular SMCs that were cultured on PGA scaffolds within bioreactors for 8 weeks. After culture, vessels were decellularized with a combination of detergents, high–ionic strength salts, and extensive washes in neutral buffers. Scanning electron microscopy showed that the acellular engineered vessels contained pores within the wall that were previously occupied by vascular SMCs (Fig. 2G). Over implant durations of up to 1 month, acellular canine and porcine vessels maintained preimplantation dimensions and resisted dilatation while stimulating very little foreign-body

reaction, as evidenced by few or absent multinucleated giant cells (*49*).

To assess longer-term remodeling and host response to human acellular engineered vessels, acellular vessels 6 mm in diameter and 20 to 30 cm in length were implanted into African baboons (*Papio anubis*), an old-world primate with strong phylogenetic proximity to *Homo sapiens*. Human acellular vessels (HAVs) had rupture strengths >3000 mmHg and suture retention strengths of >100 g before implantation. HAVs were anastomosed as arteriovenous conduits in the upper limbs of male baboons for 1 to 6 months, after which time the vessels were explanted and examined histologically (Fig. 2F) (*49*). Repopulation of the HAV walls with host cells occurred primarily from the abluminal (adventitial) surface, with cells expressing both fibroblast as well as vascular SMC markers. On the lumen, gradual repopulation of the blood-contacting surface with cells expressing von Willebrand factor (an EC marker) was evident, with repopulation occurring across the length of the vessels, implying that EC repopulation did not originate from migrating endothelium from the anastomoses and likely came from a circulating population of endothelial progenitor cells (*50*). HAVs maintained original diameter and length over implant periods of up to 6 months and demonstrated cellular repopulation of needle cannulation injuries that simulated hemodialysis access, thereby showing the durability of the acellular conduits in a high-flow in vivo environment.

Similar approaches have been taken to producing engineered, acellular heart valves (*51*). Baaijens, Hoerstrup, and colleagues cultured human fibroblasts on a PGA scaffold that was fashioned as a trileaflet valve within a tubular conduit for 4 weeks (Fig. 2H). At the end of culture, the engineered human valves were decellularized with Triton and sodium deoxycholate detergents, combined with Benzonase nuclease and buffer washes. Valves were then implanted as pulmonary heart valves into chacma baboons and followed for up to 8 weeks, after which time valves were harvested for analysis. Some evidence of retraction of the engineered leaflets was observed at the time of explant (Fig. 2I), but overall valves demonstrated good function with minimal or modest valvular regurgitation. Notably, when compared with decellularized native human heart valves in the baboon model, the cellular repopulation of the engineered valves was more rapid and complete, which may imply that cellular migration and remodeling of engineered matrix are easier than for decellularized native adult tissues.

Vascular growth

Intriguingly, acellular engineered vessels may have the capacity to grow after implantation into young recipients. Reports from Tranquillo and co-workers (*52*) have described vessels engineered from sheep-derived (or human-derived) dermal fibroblasts seeded into a fibrin matrix and cultured under pulsatile conditions in a bioreactor for 7 weeks (*53*). Decellularized conduits were implanted as pulmonary artery replacements in a juvenile dorset lamb model and were followed for nearly a year into adulthood. Lambs increased their body weight by >300% during the follow-up period, during which time the engineered vessels increased in diameter by 50% and in length by over 50% (Fig. 2, J and K). These implanted vessels maintained mechanical strength and collagen content. Combined with the extensive cellular repopulation of the implants, these results support the notion that the originally implanted acellular vessels underwent true somatic growth, as opposed to simply slow dilatation due to mechanical failure (*52*). These results provide insight that other decellularized, engineered cardiovascular tissues may also be capable of somatic growth if implanted into growing individuals, though further experimentation will be needed to prove this hypothesis.

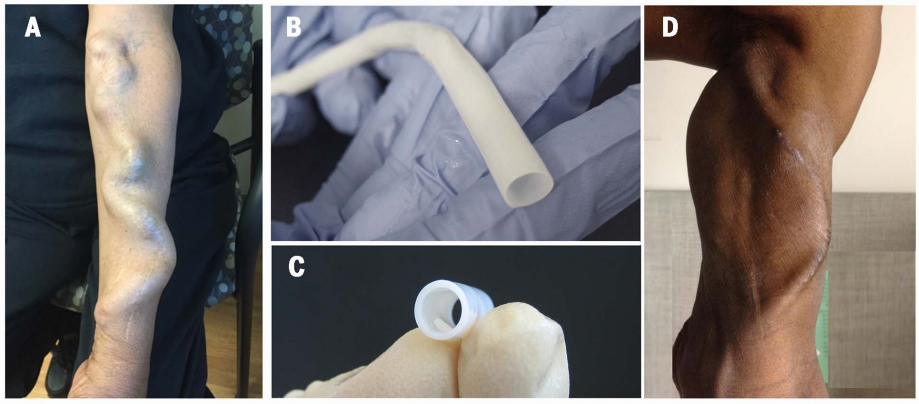


Fig. 3. Conduits for hemodialysis access. (**A**) Typical appearance of an autogenous arteriovenous fistula, showing tortuous and dilated autologous vein. (**B** and **C**) Gross photos of engineered human acellular vessels (HAVs), 6 mm in diameter. (**D**) HAV implanted as an arteriovenous conduit for dialysis access, 1 year after implantation in the upper arm of the patient.

Commercial impacts and challenges

Arterial bypass or repair for the treatment of atherosclerotic disease, infection, or traumatic injury are some of the most commonly performed surgical procedures in the Western world. In the United States alone, hundreds of thousands of coronary and peripheral arteries are bypassed, repaired, or replaced every year. Buoyed by the relative success of a few autologous approaches for vascular replacement, a number of companies have been formed to address the challenge of producing tissue-engineered blood vessels at a clinically relevant scale. The combined design and clinical criteria of nonimmunogenicity, immediate availability, and outstanding mechanical properties are extremely challenging, however, and have led to several companies initiating, and then abandoning, efforts at commercial vessel engineering.

Hence, despite the enormity of both the clinical problem and the potential market for engineered arterial replacements, the daunting challenges of immune acceptance, requisite tissue mechanics, low thrombogenicity, and immediate availability have thus far prevented engineered arterial replacements from becoming broadly available. This is in contrast to other engineered connective tissue types such as skin (which can tolerate some degree of immune recognition) or cartilage (which has an element of immune privilege). Furthermore, skin and cartilage are both tissues that may not require immediate availability and that can be implanted without having the full mechanical properties of the corresponding native tissue. Hence, because of important differences in design and functional requirements, engineered cartilage and skin products have been widely clinically available for nearly two decades, whereas engineered arteries have lagged behind.

Vascular access for hemodialysis

Availability of vascular conduits that can provide safe, infection-free, and durable access for hemodialysis patients remains a substantial challenge and an unmet clinical need. To date, modes of vascular access for hemodialysis include chronic indwelling catheters, synthetic vascular conduits and, when suitably matured, autogenous arteriovenous fistulas. Each of these existing modes of vascular access has clinical limitations. Catheters can be easily placed but offer only short-term benefits and are complicated in the longer term by infection, poor dialysis function, and the induction of central vein pathology (*54*, *55*). Synthetic grafts, made primarily of ePTFE, are easily implantable but suffer from chronic

failure because of venous intimal hyperplasia, foreign-body response, and the ongoing risk of infection (*56*). Autogenous arteriovenous fistulae are the preferred choice for vascular access owing to low rates of infection and long-term usability, but they suffer from a lack of "maturation" and function in nearly half of fistulas placed (*57*). Conversely, some autogenous fistulae can become extremely dilated (Fig. 3A), which can be cosmetically disfiguring and which can lead to vascular complications and heart failure. Hence, the development of an easily implantable, durable, and infection-resistant conduit for hemodialysis access remains critical for improving patient care.

Engineered vessels made from autologous fibroblast sheets were studied in phase 1 trials in hemodialysis access. However, the layered fibroblast-based tissues suffered some early failures and dilatation over time (*58*, *59*), and clinical development of these vessels has not progressed. More recently, the utility of HAVs in hemodialysis access has been demonstrated (*60*). Phase 1/2 studies showed that HAVs retained their diameter and flow rates, had a low incidence of infection, and had a functional patency of 89% after 1 year (Fig. 3, B to D). Over the long term, some patients have continued using the HAV for dialysis access after more than 6 years. After implantation into patients, HAVs also undergo a gradual process of host cell repopulation, which may contribute to their durability and infection resistance over time (see Fig. 4). Host cell repopulation also facilitates healing of cannulation sites from thrice-weekly needle punctures for hemodialysis (*43*, *60*). Phase 3 trials of the HAV in hemodialysis are currently under way.

It is important to note that HAVs may have some potential drawbacks. The vessels are far more expensive to manufacture than are ePTFE grafts and will remain so, despite reductions in cost-of-goods that will certainly occur over time. In addition, the incidence of pseudoaneurysm formation in HAVs—which is the leakage of blood from needle holes in the conduit after completion of dialysis—appeared to be somewhat higher than has been reported for ePTFE grafts. This observation may mean that additional training of access cannulation personnel and interventionalists for this new technology may be required for optimal clinical function and utility.

Peripheral arterial disease and vascular trauma

The need for an immediately available "conduit" for bypass in both peripheral arterial disease (PAD) and vascular trauma remains acute. Previously, no tissue-engineered vascular construct has demonstrated utility in these challenging clinical applications; however, clinical studies in both PAD and vascular trauma are currently under way (Fig. 4). There are now more than 60 human implants in these clinical settings, with follow-up times of up to 6 years in some patients. HAVs have demonstrated long-term durability (Fig. 4, left panels) (*61*), limb salvage in the setting of trauma (*62*), and an avoidance of infection.

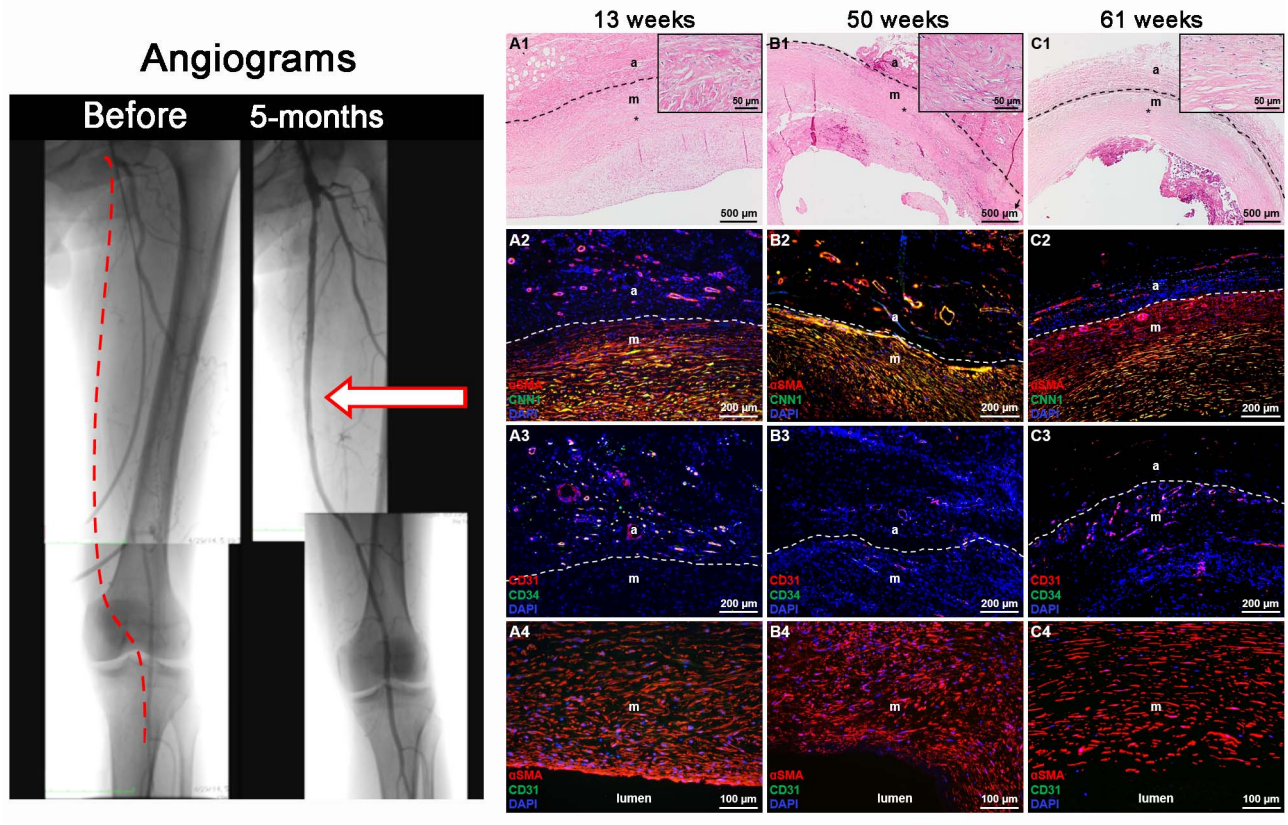


Fig. 4. Peripheral arterial bypass using HAVs. Left panels: Angiograms showing preoperative and postoperative (5 months) arterial blood flow in the operative leg of a patient having occlusive arterial disease, who received a bypass using an HAV (arrow). Right panels: Microscopic evaluation of three explant samples from patients treated with HAVs for peripheral arterial disease (*61*). In (A1) to (C4), "m" indicates media, or the wall of the HAV, while "a" denotes adventitia, or the surrounding area outside of the HAV. (A1 to C1) H&E images of samples explanted at 13, 50, and 61 weeks, respectively. Scale bars, 500 μm. (A2 to C2) Immunofluorescence staining for alpha smooth muscle actin (αSMA, red), calponin (CNN1, green), and 4′,6-diamidino-2-phenylindole (DAPI) for nuclei (blue). (A3 to C3) Immunofluorescence for PECAM-1 (CD31, red), CD34 (green), and DAPI shows microvascular endothelium in adventitia and outer vessel wall. (A4 to C4) Immunofluorescence for actin (αSMA, red), CD31 (green), and DAPI shows extensive αSMA expression near lumen at all time points.

Although still representing a "first in human" experience, these surgical cases occurred under conditions in which no natural venous conduit was available and where the use of ePTFE was unfavorable or at high risk for infection.

In clinical research subjects, HAVs implanted in the arterial position gradually repopulate with spindle-shaped host cells (Fig. 4, A1 to C1). Cells repopulating the "media" of the engineered vessel express both α-smooth muscle actin and calponin (Fig. 4, A2 to C2), implying a vascular smooth muscle phenotype of the host cells invading the vessel wall. Microvessels expressing the progenitor cell marker CD34 develop in the neo-adventitia at early time points after implantation (Fig. 4A3), whereas at later time points, microvessels are visible in the repopulated wall and express CD31, an endothelial marker (Fig. 4, B3 and C3). Although endothelium expressing CD31 is not clearly demonstrated in the lumen of these explant samples, other samples have demonstrated reendothelialization of the HAV lumens over time (*43*, *49*, *60*).

Challenges of scaling human tissue production

For a first-in-class technology to have true clinical impact, it must be manufacturable at scales that permit wide dissemination at reasonable cost. Challenges of scaling any biological process include maintaining cellular phenotype across extended passaging and culture times, controlling uniformity and reproducibility of product, and ensuring reliably sterile conditions for tissue growth. The industrialization of cell-based manufacturing began in the 1990s with the production of therapeutic proteins such as erythropoietin. Initially, biotechnology companies had to surmount many obstacles while developing reliable, robust biomanufacturing processes having high yields. Although the production of cell-based protein therapeutics is now almost commonplace, several decades of technological development were required before procedures and best practices became standardized.

Similar work is now under way with the production of functional human tissues that are derived from allogeneic cells. A small number of entities have worked successfully to "industrialize" the culture of human tissues and standardized processes for growing human skin replacements (*63*). Automated technologies have recently been developed to produce acellular engineered human arteries at scale (Fig. 5). Fundamental to any scaled and controlled process is focused attention on the control of key culture conditions such as oxygen tension, pH, nutrients and waste products, and culture temperature. Combined with environmental and culture controls, tracking and controlling cell phenotype throughout the culture process is essential to reliable success. Lastly, automation of repetitive steps is crucial to minimizing human errors during extended periods of engineered tissue culture (Fig. 5). Notably, lessons learned from the automated culture of one tissue type are typically transferrable to the culture of other types of tissues, because the underlying principles of environmental and metabolic controls are fairly universal.

Alternative approaches and potential future directions

Although acellular engineered constructs seem to hold great promise for the time being, alternative technologies may one day displace this approach. Many investigators are now evaluating the capabilities of electrospinning of synthetic and natural materials as a means of producing tubular constructs having controlled mechanical properties and compositions (Fig. 6, A to D) (*64*, *65*). Although initially developed as a material processing technique for the textiles industry in the 1930s, electrospinning has seen a resurgence in recent years in the bioengineering community, with the production of vascular grafts derived from many different materials (*66*, *67*). The pore size of electrospun materials can often be a hindrance to cellular infiltration (Fig. 6A); however, the relative control over mass density and geometry, as well as fiber orientation, may provide new tools that may result in a superior vascular graft in the future.

Similarly, "bioprinting" of cells to form vascular conduits has been reported over the past 10 years. One vascular prototype made from spheroids of SMCs or fibroblasts was reported by Norotte and colleagues in 2009 (Fig. 6E) (*68*). Similar spheroid-based approaches have been used by other investigators to bioprint cardiac tissue tubes that demonstrate some degree of contractility (Fig. 6, F and G) (*69*). Lastly, very recent advances in bioprinting using photopolymerizable hydrogels combined with projection stereolithography have produced complex microvascular channels that mimic some aspects of organ microvasculature (Fig. 6H) (*70*). Although these bioprinted structures have not yet demonstrated sufficient mechanical integrity for in vivo implantation, owing to a lack of cross-linked mature extracellular matrix proteins, it is conceivable that bioprinted, mature and cross-linked fibrillar collagen may one day be possible, which could revolutionize the engineering of connective tissues and solid organs.

Recent advances in collagen bioprinting have brought us somewhat closer to functional vascular constructs. Lee and colleagues recently described the "freeform reversible embedding of suspended hydrogels" ("FRESH") method, which allows the extrusion of acid-solubilized collagen into a bath of controlled temperature and pH that drives rapid collagen self-assembly (*71*). Complex structures, including those mimicking a human neonatal heart, valves, and blood vessels, have been fabricated. Although no mechanical characterization was reported for these structures, constructs were fixed in either ethanol or formaldehyde before functional testing, implying that the collagen gels as printed were quite soft. Collagen gels typically have moduli on the order of 1 to 5 kPa (*72*) which is far below the 20 to 100 MPa reported for native collagen fibers. Despite exciting and rapid progress, it is currently not possible to bioprint mechanically robust collagen molecules comprising physiological cross-links that can be remodeled by mammalian cells.

Conclusions

Extraordinary progress has been made in recent decades toward the development of engineered vasculature. Beginning with silicone-based materials and evolving into the industrialization of vessel production, iterative improvements have been driven by the biology and physiology of the cardiovascular system. By attending to key design principles for successful arterial replacement, the field of vascular

Fig. 5. Automated system for engineered vessel production. Automated and self-contained system for the culture and decellularization of engineered human blood vessels.

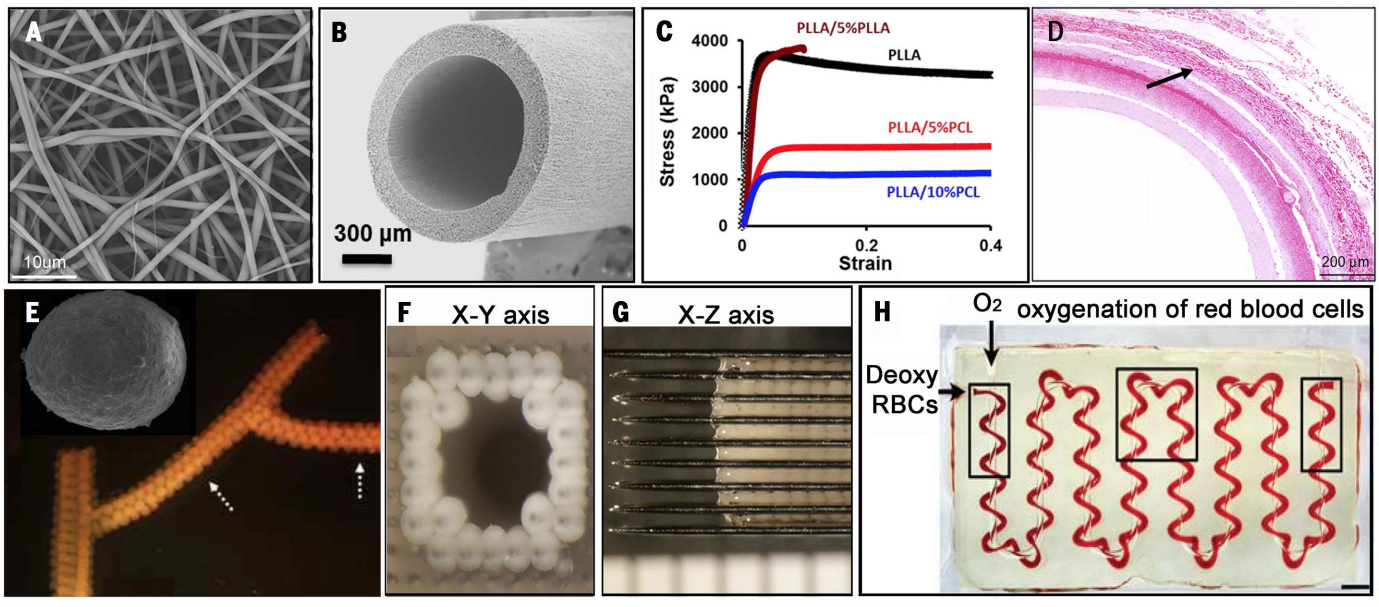


Fig. 6. Possible future technologies for cardiovascular engineering. (**A** to **D**) Electrospinning. (A) Electrospun polymer fibers shown under scanning electron microscopy (SEM) (*74*). (B) SEM image of cut end of electrospun conduit. (C) Stress-strain curves of electrospun vessels (*64*). (D) H&E image of cellular repopulation outside of electrospun scaffolds after 3 months in a rabbit carotid artery implantation position [arrow indicates fibroblasts; modified from (*74*)]. (**E** to **H**) Bioprinting. (E) Bioprinted human smooth muscle cell sphere and printed branching vessel structure, immediately after printing [dotted arrows indicate 0.9-mm-diameter; modified from (*68*)]. (F and G) Bioprinted cardiac tissue, with cell spheroids positioned using needles [modified from (*69*)]. (H) Bioprinted vascular structure using optically based sacrificial hydrogel framework, showing oxygenation of red blood cells (RBCs) during left-to-right flow through the network [modified from (*70*)].

engineering is making great strides. Therapeutic engineered tissues for arterial and microvascular diseases are becoming a clinical reality, largely due to the convergence of biological, engineering, and design principles.

REFERENCES AND NOTES

1. MIT, *Remembering Eugene Bell* (15 October 2007); www.technologyreview.com/s/408864/remembering-eugene-bell/.
2. R. L. Armentano *et al*., Assessment of elastin and collagen contribution to aortic elasticity in conscious dogs. *Am. J. Physiol.* **260**, H1870–H1877 (1991). pmid: 1905490
3. CDC, *Data and Statistics on Congenital Heart Defects* (2019); www.cdc.gov/ncbddd/heartdefects/data.html.
4. W. A. Lane, The award of the Nobel Prize to Dr. Alexis Carrel. *Lancet* **180**, 1103 (1912). doi: 10.1016/S0140-6736(01)57384-5
5. A. Carrel, C. H. Lindbergh, *The Culture of Organs* (Hoeber, New York, 1938).
6. C. H. Sparks, Development of a successful silicone rubber arterial graft. *Ann. Thorac. Surg.* **2**, 585–593 (1966). doi: 10.1016/S0003-4975(10)66620-9; pmid: 5934073
7. C. H. Sparks, Die-grown reinforced arterial grafts: Observations on long-term animal grafts and clinical experience. *Ann. Surg.* **172**, 787–794 (1970). doi: 10.1097/00000658-197011000-00003; pmid: 5477654
8. R. W. Hallin, Complications with the mandril-grown (Sparks) dacron arterial graft. *Am. Surg.* **41**, 550–554 (1975). pmid: 126653
9. R. W. Hallin, W. R. Sweetman, The Sparks' mandril graft. A seven year follow-up of mandril grafts placed by Charles H. Sparks and his associates. *Am. J. Surg.* **132**, 221–223 (1976). doi: 10.1016/0002-9610(76)90051-9; pmid: 133619
10. W. J. Geelhoed, L. Moroni, J. I. Rotmans, Utilizing the foreign body response to grow tissue engineered blood vessels in vivo. *J. Cardiovasc. Transl. Res.* **10**, 167–179 (2017). doi: 10.1007/s12265-017-9731-7; pmid: 28205013
11. C. B. Weinberg, E. Bell, A blood vessel model constructed from collagen and cultured vascular cells. *Science* **231**, 397–400 (1986). doi: 10.1126/science.2934816; pmid: 2934816
12. P. Zilla *et al*., Clinical in vitro endothelialization of femoropopliteal bypass grafts: An actuarial follow-up over three years. *J. Vasc. Surg.* **19**, 540–548 (1994). doi: 10.1016/S0741-5214(94)70083-4; pmid: 8126869
13. M. Deutsch, J. Meinhart, T. Fischlein, P. Preiss, P. Zilla, Clinical autologous in vitro endothelialization of infrainguinal ePTFE grafts in 100 patients: A 9-year experience. *Surgery* **126**, 847–855 (1999). doi: 10.1016/S0039-6060(99)70025-5; pmid: 10568184
14. M. Deutsch *et al*., Long-term experience in autologous in vitro endothelialization of infrainguinal ePTFE grafts. *J. Vasc. Surg.* **49**, 352–362, discussion 362 (2009). doi: 10.1016/j.jvs.2008.08.101; pmid: 19110397
15. N. L'Heureux, S. Pâquet, R. Labbé, L. Germain, F. A. Auger, A completely biological tissue-engineered human blood vessel. *FASEB J.* **12**, 47–56 (1998). pmid: 9438410
16. L. E. Niklason, R. S. Langer, Advances in tissue engineering of blood vessels and other tissues. *Transpl. Immunol.* **5**, 303–306 (1997). doi: 10.1016/S0966-3274(97)80013-5; pmid: 9504152
17. L. E. Niklason *et al*., Morphologic and mechanical characteristics of engineered bovine arteries. *J. Vasc. Surg.* **33**, 628–638 (2001). doi: 10.1067/mva.2001.111747; pmid: 11241137
18. L. E. Niklason *et al*., Functional arteries grown in vitro. *Science* **284**, 489–493 (1999). doi: 10.1126/science.284.5413.489; pmid: 10205057
19. D. Seliktar, R. A. Black, R. P. Vito, R. M. Nerem, Dynamic mechanical conditioning of collagen-gel blood vessel constructs induces remodeling in vitro. *Ann. Biomed. Eng.* **28**, 351–362 (2000). doi: 10.1114/1.275; pmid: 10870892
20. D. Seliktar, R. M. Nerem, Z. S. Galis, Mechanical strain-stimulated remodeling of tissue-engineered blood vessel constructs. *Tissue Eng.* **9**, 657–666 (2003). doi: 10.1089/10763270376824735 9; pmid: 13678444
21. L. E. Niklason, Replacement arteries made to order. *Science* **286**, 1493–1494 (1999). doi: 10.1126/science.286.5444.1493; pmid: 10610551
22. D. Y. Li *et al*., Novel arterial pathology in mice and humans hemizygous for elastin. *J. Clin. Invest.* **102**, 1783–1787 (1998). doi: 10.1172/JCI4487; pmid: 9819363
23. A. H. Huang *et al*., Biaxial stretch improves elastic fiber maturation, collagen arrangement, and mechanical properties in engineered arteries. *Tissue Eng. Part C Methods* **22**, 524–533 (2016). doi: 10.1089/ten.tec.2015.0309; pmid: 27108525
24. J. L. Long, R. T. Tranquillo, Elastic fiber production in cardiovascular tissue-equivalents. *Matrix Biol.* **22**, 339–350 (2003). doi: 10.1016/S0945-053X(03)00052-0; pmid: 12935818
25. J. Y. Liu *et al*., Functional tissue-engineered blood vessels from bone marrow progenitor cells. *Cardiovasc. Res.* **75**, 618–628 (2007). doi: 10.1016/j.cardiores.2007.04.018; pmid: 17512920
26. A. Fernández-Colino *et al*., Small caliber compliant vascular grafts based on elastin-like recombinamers for in situ tissue engineering. *Front. Bioeng. Biotechnol.* **7**, 340 (2019). doi: 10.3389/fbioe.2019.00340; pmid: 31803735
27. T. Shin'oka, Y. Imai, Y. Ikada, Transplantation of a tissue-engineered pulmonary artery. *N. Engl. J. Med.* **344**, 532–533 (2001). doi: 10.1056/NEJM200102153440717; pmid: 11221621
28. J. D. Drews, H. Miyachi, T. Shinoka, Tissue-engineered vascular grafts for congenital cardiac disease: Clinical experience and current status. *Trends Cardiovasc. Med.* **27**, 521–531 (2017). doi: 10.1016/j.tcm.2017.06.013; pmid: 28754230
29. T. Shoji, T. Shinoka, Tissue engineered vascular grafts for pediatric cardiac surgery. *Transl. Pediatr.* **7**, 188–195 (2018). doi: 10.21037/tp.2018.02.01; pmid: 29770300
30. NIH, Clinical Trials.Gov 2019 October 20th 2019; https://clinicaltrials.gov/ct2/show/NCT01034007?term=shinoka&draw=2&rank=1.
31. P. T. Kennealey *et al*., A prospective, randomized comparison of bovine carotid artery and expanded polytetrafluoroethylene for permanent hemodialysis vascular access. *J. Vasc. Surg.* **53**, 1640–1648 (2011). doi: 10.1016/j.jvs.2011.02.008; pmid: 21609797
32. H. E. Katzman, M. H. Glickman, A. F. Schild, R. M. Fujitani, J. H. Lawson, Multicenter evaluation of the bovine mesenteric vein bioprostheses for hemodialysis access in patients with an earlier failed prosthetic graft. *J. Am. Coll. Surg.* **201**, 223–230 (2005). doi: 10.1016/j.jamcollsurg.2005.03.040; pmid: 16038820
33. S. Manduz, N. Katrancioglu, E. Ozker, K. Dogan, Early thrombosis in bovine mesenteric vein grafts after infrainguinal reconstruction. *Int. J. Angiol.* **17**, 37–39 (2008). doi: 10.1055/s-0031-1278278; pmid: 22477370
34. M. T. Koobatian *et al*., Successful endothelialization and remodeling of a cell-free small-diameter arterial graft in a large animal model. *Biomaterials* **76**, 344–358 (2016). doi: 10.1016/j.biomaterials.2015.10.020; pmid: 26561932
35. H. Peng, E. M. Schlaich, S. Row, S. T. Andreadis, D. D. Swartz, A novel ovine ex vivo arteriovenous shunt model to test

vascular implantability. *Cells Tissues Organs* **195**, 108–121 (2012). doi: 10.1159/000331415; pmid: 22005667
36. N. Dobrilovic, P. Soukas, I. Sadiq, L. Goldstein, J. Raman, Early complications of biologic extracellular matrix patch after use for femoral artery repair. *J. Vasc. Surg.* **65**, 705–710 (2017). doi: 10.1016/j.jvs.2016.07.131; pmid: 27751739
37. M. Lopez-Cepero, C. E. Sanders, J. Buggs, V. Bowers, Sensitization of renal transplant candidates by cryopreserved cadaveric venous or arterial allografts. *Transplantation* **73**, 817–819 (2002). doi: 10.1097/00007890-200203150-00029; pmid: 11907435
38. B. Benedetto *et al*., Use of cryopreserved cadaveric vein allograft for hemodialysis access precludes kidney transplantation because of allosensitization. *J. Vasc. Surg.* **34**, 139–142 (2001). doi: 10.1067/mva.2001.114206; pmid: 11436087
39. J. Merola, D. D. Jane-Wit, J. S. Pober, Recent advances in allograft vasculopathy. *Curr. Opin. Organ Transplant.* **22**, 1–7 (2017). doi: 10.1097/MOT.0000000000000370; pmid: 27898462
40. J. S. Pober, D. Jane-wit, L. Qin, G. Tellides, Interacting mechanisms in the pathogenesis of cardiac allograft vasculopathy. *Arterioscler. Thromb. Vasc. Biol.* **34**, 1609–1614 (2014). doi: 10.1161/ATVBAHA.114.302818; pmid: 24903097
41. J. G. Barra *et al*., Assessment of smooth muscle contribution to descending thoracic aortic elastic mechanics in conscious dogs. *Circ. Res.* **73**, 1040–1050 (1993). doi: 10.1161/01.RES.73.6.1040; pmid: 8222076
42. D. W. Courtman, B. F. Errett, G. J. Wilson, The role of crosslinking in modification of the immune response elicited against xenogenic vascular acellular matrices. *J. Biomed. Mater. Res.* **55**, 576–586 (2001). doi: 10.1002/1097-4636(20010615)55:4<576::AID-JBM1051>3.0.CO;2-9; pmid: 11288086
43. R. D. Kirkton *et al*., Bioengineered human acellular vessels recellularize and evolve into living blood vessels after human implantation. *Sci. Transl. Med.* **11**, eaau6934 (2019). doi: 10.1126/scitranslmed.aau6934; pmid: 30918113
44. H. He, T. Shirota, H. Yasui, T. Matsuda, Canine endothelial progenitor cell-lined hybrid vascular graft with nonthrombogenic potential. *J. Thorac. Cardiovasc. Surg.* **126**, 455–464 (2003). doi: 10.1016/S0022-5223(02)73264-9; pmid: 12928644
45. R. L. Madden, G. S. Lipkowitz, B. J. Browne, A. Kurbanov, A comparison of cryopreserved vein allografts and prosthetic grafts for hemodialysis access. *Ann. Vasc. Surg.* **19**, 686–691 (2005). doi: 10.1007/s10016-005-6621-2; pmid: 16075346
46. R. Madden *et al*., Decellularized cadaver vein allografts used for hemodialysis access do not cause allosensitization or preclude kidney transplantation. *Am. J. Kidney Dis.* **40**, 1240–1243 (2002). doi: 10.1053/ajkd.2002.36892; pmid: 12460043
47. C. B. Basu, L. Jeffers, The role of acellular dermal matrices in capsular contracture: A review of the evidence. *Plast. Reconstr. Surg.* **130** (suppl. 2), 118S–124S (2012). doi: 10.1097/PRS.0b013e318262df58; pmid: 23096960
48. S. L. M. Dahl, J. Koh, V. Prabhakar, L. E. Niklason, Decellularized native and engineered arterial scaffolds for transplantation. *Cell Transplant.* **12**, 659–666 (2003). doi: 10.3727/000000003108747136; pmid: 28866947
49. S. L. Dahl *et al*., Readily available tissue-engineered vascular grafts. *Sci. Transl. Med.* **3**, 68ra9 (2011). doi: 10.1126/scitranslmed.3001426; pmid: 21289273
50. A. W. Clowes, A. M. Gown, S. R. Hanson, M. A. Reidy, Mechanisms of arterial graft failure. 1. Role of cellular proliferation in early healing of PTFE prostheses. *Am. J. Pathol.* **118**, 43–54 (1985). pmid: 3966536
51. B. Weber *et al*., Off-the-shelf human decellularized tissue-engineered heart valves in a non-human primate model. *Biomaterials* **34**, 7269–7280 (2013). doi: 10.1016/j.biomaterials.2013.04.059; pmid: 23810254
52. Z. Syedain *et al*., Tissue engineering of acellular vascular grafts capable of somatic growth in young lambs. *Nat. Commun.* **7**, 12951 (2016). doi: 10.1038/ncomms12951; pmid: 27676438
53. Z. H. Syedain *et al*., A completely biological "off-the-shelf" arteriovenous graft that recellularizes in baboons. *Sci. Transl. Med.* **9**, eaan4209 (2017). doi: 10.1126/scitranslmed.aan4209; pmid: 29093182
54. A. K. Agarwal, N. Haddad, K. Boubes, Avoiding problems in tunneled dialysis catheter placement. *Semin. Dial.* **32**, 535–540 (2019). doi: 10.1111/sdi.12845; pmid: 31710156
55. L. Kumbar, J. Yee, Current concepts in hemodialysis vascular access infections. *Adv. Chronic Kidney Dis.* **26**, 16–22 (2019). doi: 10.1053/j.ackd.2018.10.005; pmid: 30876612
56. U. Hedin, Long-term results of PTFE grafts. *J. Vasc. Access* **16** (suppl. 9), S87–S92 (2015). doi: 10.5301/jva.5000350; pmid: 25751559
57. R. L. Pisoni *et al*., International differences in the location and use of arteriovenous accesses created for hemodialysis: Results from the dialysis outcomes and practice patterns study (DOPPS). *Am. J. Kidney Dis.* **71**, 469–478 (2018). doi: 10.1053/j.ajkd.2017.09.012; pmid: 29198387
58. T. N. McAllister *et al*., Effectiveness of haemodialysis access with an autologous tissue-engineered vascular graft: A multicentre cohort study. *Lancet* **373**, 1440–1446 (2009). doi: 10.1016/S0140-6736(09)60248-8; pmid: 19394535
59. W. Wystrychowski *et al*., Case study: First implantation of a frozen, devitalized tissue-engineered vascular graft for urgent hemodialysis access. *J. Vasc. Access* **12**, 67–70 (2011). doi: 10.5301/JVA.2011.6360; pmid: 21360466
60. J. H. Lawson *et al*., Bioengineered human acellular vessels for dialysis access in patients with end-stage renal disease: Two phase 2 single-arm trials. *Lancet* **387**, 2026–2034 (2016). doi: 10.1016/S0140-6736(16)00557-2; pmid: 27203778
61. P. Gutowski *et al*., Arterial reconstruction with human bioengineered acellular blood vessels in patients with peripheral arterial disease. *J. Vasc. Surg.* **72**, 1247–1258 (2020). pmid: 32093913
62. J. J. Morrison *et al*., Clinical implementation of the Humacyte human acellular vessel: Implications for military and civilian trauma care. *J. Trauma Acute Care Surg.* **87** (suppl. 1), S44–S47 (2019). doi: 10.1097/TA.0000000000002350; pmid: 31246906
63. L. Zaulyanov, R. S. Kirsner, A review of a bi-layered living cell treatment (Apligraf) in the treatment of venous leg ulcers and diabetic foot ulcers. *Clin. Interv. Aging* **2**, 93–98 (2007). doi: 10.2147/ciia.2007.2.1.93; pmid: 18044080
64. J. J. D. Henry *et al*., Engineering the mechanical and biological properties of nanofibrous vascular grafts for in situ vascular tissue engineering. *Biofabrication* **9**, 035007 (2017). doi: 10.1088/1758-5090/aa834b; pmid: 28817384
65. R. Khosravi *et al*., Long-term functional efficacy of a novel electrospun poly(glycerol sebacate)-based arterial graft in mice. *Ann. Biomed. Eng.* **44**, 2402–2416 (2016). doi: 10.1007/s10439-015-1545-7; pmid: 26795977
66. X. Qiu *et al*., End-point immobilization of heparin on plasma-treated surface of electrospun polycarbonate-urethane vascular graft. *Acta Biomater.* **51**, 138–147 (2017). doi: 10.1016/j.actbio.2017.01.012; pmid: 28069505
67. T. Sultana, J. Amirian, C. Park, S. J. Lee, B.-T. Lee, Preparation and characterization of polycaprolactone-polyethylene glycol methyl ether and polycaprolactone-chitosan electrospun mats potential for vascular tissue engineering. *J. Biomater. Appl.* **32**, 648–662 (2017). doi: 10.1177/0885328217733849; pmid: 28956678
68. C. Norotte, F. S. Marga, L. E. Niklason, G. Forgacs, Scaffold-free vascular tissue engineering using bioprinting. *Biomaterials* **30**, 5910–5917 (2009). doi: 10.1016/j.biomaterials.2009.06.034; pmid: 19664819
69. K. Arai *et al*., Fabrication of scaffold-free tubular cardiac constructs using a Bio-3D printer. *PLOS ONE* **13**, e0209162 (2018). doi: 10.1371/journal.pone.0209162; pmid: 30557409
70. B. Grigoryan *et al*., Multivascular networks and functional intravascular topologies within biocompatible hydrogels. *Science* **364**, 458–464 (2019). doi: 10.1126/science.aav9750; pmid: 31048486
71. A. Lee *et al*., 3D bioprinting of collagen to rebuild components of the human heart. *Science* **365**, 482–487 (2019). doi: 10.1126/science.aav9051; pmid: 31371612
72. I. D. Gaudet, D. I. Shreiber, Characterization of methacrylated type-I collagen as a dynamic, photoactive hydrogel. *Biointerphases* **7**, 25 (2012). doi: 10.1007/s13758-012-0025-y; pmid: 22589068
73. R. F. Rickard, D. A. Hudson, A history of vascular and microvascular surgery. *Ann. Plast. Surg.* **73**, 465–472 (2014). doi: 10.1097/SAP.0b013e3182710027; pmid: 23318371
74. D. Jin *et al*., Evaluation of a simple off-the-shelf bi-layered vascular scaffold based on poly(L-lactide-co-ε-caprolactone)/silk fibroin in vitro and in vivo. *Int. J. Nanomedicine* **14**, 4261–4276 (2019). doi: 10.2147/IJN.S205569; pmid: 31289441

ACKNOWLEDGMENTS

We thank H. Prichard, R. Kirkton, and R. Schutte for assistance with figures and content. **Competing interests:** L.E.N. and J.H.L. are shareholders of Humacyte Inc., a regenerative medicine company that develops engineered human blood vessels.

10.1126/science.aaw8682

REVIEW SUMMARY

ELECTROCHEMISTRY

Recent advances in solid oxide cell technology for electrolysis

A. Hauch*†, R. Küngas†, P. Blennow‡, A. B. Hansen‡, J. B. Hansen‡, B. V. Mathiesen‡, M. B. Mogensen‡

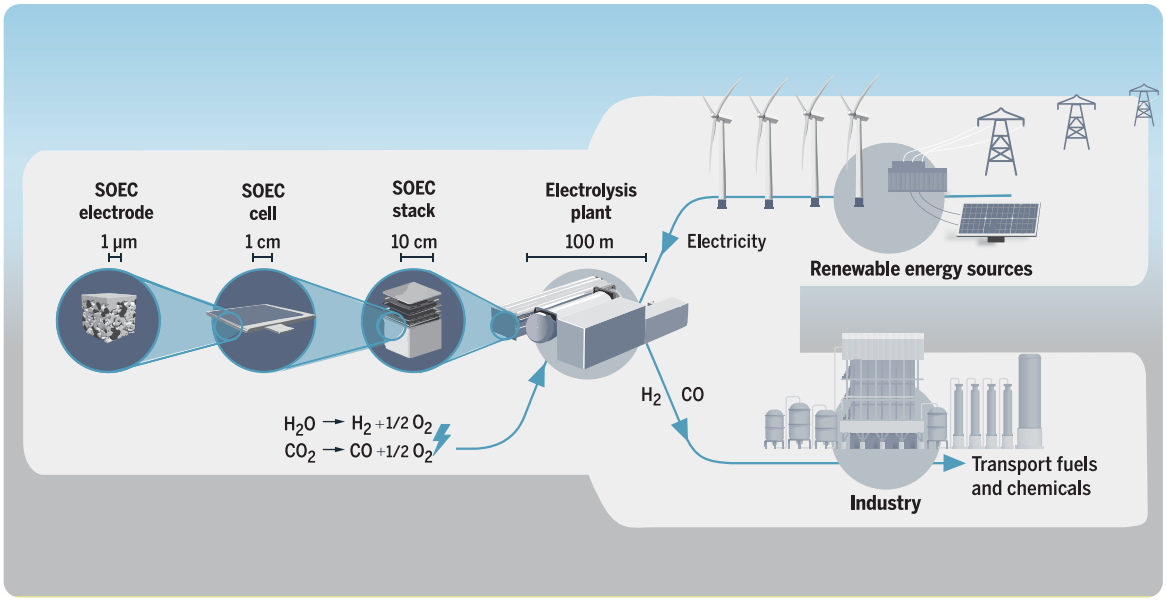
BACKGROUND: Alleviating the worst effects of climate change requires drastic modification of our energy system: moving from fossil fuels to low-carbon energy sources. The challenge is not the amount of renewable energy available—energy potential from solar and wind exceeds global energy consumption many times over. Rather, the key to a 100% renewable energy supply lies in the integration of the growing share of intermittent sources into a power infrastructure that can meet continuous demand. The higher the share of renewables, the more flexible and interconnected the energy system (the electric grid, the gas and heat networks, etc.) needs to be. Critically, a future energy system where the supply of electricity, heat, and fuels is based solely on renewables relies heavily on technologies capable of converting electricity into chemicals and fuels suitable for heavy transport at high efficiencies. In addition, higher electrolysis efficiency and integrated fuel production can decrease the reliance on bioenergy further than conventional electrolysis can.

ADVANCES: Electrolysis is the core technology of power-to-X (PtX) solutions, where X can be hydrogen, syngas, or synthetic fuels. When electrolysis is combined with renewable electricity, the production of fuels and chemicals can be decoupled from fossil resources, paving the way for an energy system based on 100% renewable energy. Solid oxide electrolysis cell (SOEC) technology is attractive because of unrivaled conversion efficiencies—a result of favorable thermodynamics and kinetics at higher operating temperatures. SOECs can be used for direct electrochemical conversion of steam (H_2O), carbon dioxide (CO_2), or both into hydrogen (H_2), carbon monoxide (CO), or syngas (H_2+CO), respectively. SOECs can be thermally integrated with a range of chemical syntheses, enabling recycling of captured CO_2 and H_2O into synthetic natural gas or gasoline, methanol, or ammonia, resulting in further efficiency improvements compared with low-temperature electrolysis technologies. SOEC technology has undergone tremendous development and improvements over the past 10 to 15 years. The initial electrochemical performance of state-of-the-art SOEC single cells has more than doubled, while long-term durability has been improved by a factor of ~100. Similar improvements in performance and durability have been achieved on the stack level. Furthermore, SOEC technology is based on scalable production methods and abundant raw materials such as nickel, zirconia, and steel, not precious metals. Performance and durability improvements as well as increased scale-up efforts have led to a hundredfold gas production capacity increase within the past decade and to commissioning of the first industrially relevant SOEC plants. Over the next 2 to 3 years, plant size is expected to further increase by a factor of almost 20. In recent years, SOEC systems have been integrated with downstream synthesis processes: examples include a demonstration plant for upgrading of biogas to pipeline quality methane and the use of syngas from an SOEC plant to produce fuels for transport via the Fischer-Tropsch process.

OUTLOOK: Improved understanding of the nanoscale processes occurring in SOECs will continue to result in performance and lifetime gains on the cell, stack, and system levels, which in turn will enable even larger and more efficient SOEC plants. In Germany, the share of intermittent renewables in the electricity supply has passed 30%, while in Denmark, intermittent sources account for almost 50% of the electricity supply. As this happens for a growing number of countries, demand for efficient energy conversion technologies such as SOECs is poised to increase. The increasing scale will help bring down production costs, thereby making SOECs cost-competitive with other electrolysis technologies and, given sufficiently high CO_2 emissions taxation, cost-competitive with fossil-based methods for producing H_2 and CO. SOECs offer an opportunity to decrease the costs of future renewable energy systems through more efficient conversion and enable further integration of renewables into the energy mix. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: hauc@dtu.dk
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as A. Hauch *et al.*, *Science* **370**, eaba6118 (2020). DOI: 10.1126/science.aba6118

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba6118



Solid oxide electrolyzers: From nanoscale to macroscale. The splitting of H_2O or CO_2 occurs at solid oxide electrolysis cell (SOEC) electrodes. Multiple cells are combined into SOEC stacks, which are in turn combined into SOEC plants. When renewable electricity is used, the production of transport fuels and chemicals can be decoupled from fossil resources. SOECs operate at elevated temperatures, resulting in electrolysis efficiencies unattainable by other electrolysis technologies.

REVIEW

ELECTROCHEMISTRY

Recent advances in solid oxide cell technology for electrolysis

A. Hauch[1]*†, R. Küngas[2]†, P. Blennow[2]‡, A. B. Hansen[3]‡, J. B. Hansen[2]‡, B. V. Mathiesen[4]‡, M. B. Mogensen[1]‡

In a world powered by intermittent renewable energy, electrolyzers will play a central role in converting electrical energy into chemical energy, thereby decoupling the production of transport fuels and chemicals from today's fossil resources and decreasing the reliance on bioenergy. Solid oxide electrolysis cells (SOECs) offer two major advantages over alternative electrolysis technologies. First, their high operating temperatures result in favorable thermodynamics and reaction kinetics, enabling unrivaled conversion efficiencies. Second, SOECs can be thermally integrated with downstream chemical syntheses, such as the production of methanol, dimethyl ether, synthetic fuels, or ammonia. SOEC technology has witnessed tremendous improvements during the past 10 to 15 years and is approaching maturity, driven by advances at the cell, stack, and system levels.

Our modern society relies heavily on fossil energy sources such as coal, natural gas, and petroleum. Fossil fuels still serve as a vast and inexpensive source for on-demand heat and electricity, fueling every part of the economy, from transportation to computing to heavy industry, but the share of renewables is steadily rising. However, a society based solely on low-carbon energy sources is not only possible (*1*–*3*) but also necessary to avoid the worst effects of climate change. Wind and solar could easily provide enough power to cover the global energy demand many times over (*4*), but the intermittent nature of these energy sources will require us to rethink and redesign our energy system.

Several challenges exist. First, technologies for storage and conversion of energy will have to be implemented at a massive scale to make seasonal energy storage possible. Second, the different branches of the energy system (electric grid, gas and heat networks, and transportation sector) will need to become much better integrated (*1*, *5*), allowing temporary excesses in one branch to be synergistically absorbed by other branches. Third, although it is a good idea to electrify light-duty vehicles, synthetic fuels, with their high energy density, offer important advantages in the aviation, maritime, and heavy-duty vehicle transport sectors. Fourth, the production of chemicals and commodities, such as steel, glass, and cement, must be decoupled from the use of fossil resources.

The above challenges can be addressed by a "smart" energy system, such as the one depicted in Fig. 1. Here, electricity from renewable sources is used to power electrolysis plants capable of splitting steam (H_2O) and carbon dioxide (CO_2). The resulting hydrogen (H_2) or carbon monoxide (CO) serve as feedstock for the chemical industry and fuel synthesis plants, thereby enabling various power-to-X scenarios, where X stands for any chemical or fuel. The resulting green synthetic fuels can be used to power heavy or long-haul transport (planes, ships, and trucks), while light-duty transport is powered by electricity stored in batteries. The heat demand of buildings is preferably met by district heating, for example, by heat pumps or by excess heat generated by the industry.

Considering an energy system like the one outlined in Fig. 1 raises some obvious questions: At what percentage of penetration of wind and solar in the electricity system will electrolysis be advantageous? How large a share of electricity demand can solar and wind cover? When will fluctuating renewable energy meet peak electricity demands, and how much of the time will an electrolysis plant actually be in operation? If we are to meet our need for storage of renewable energy, how many megawatts of electrolysis plants and how many terawatt-hours of chemical storage will be needed? To answer and exemplify, studies show that greater than 40 to 50% penetration of wind power and photovoltaics in the electricity system requires further sector integration in combination with efficient energy conversion and storage technologies such as electrolysis are needed (*1*, *5*). European-level analyses (as well as German- and Danish-scale studies) show that wind power and photovoltaics can cover 80 to 90% of the electricity demands in energy systems like the one sketched in Fig. 1 (*1*–*3*, *6*, *7*). Fluctuating renewable energy will meet peak electricity demands at 40 to 50% penetration even with high electrification rates (*3*), and electrolysis can thereby substantially increase the share of renewables by enabling chemical storage and by decarbonizing the transport sector. Although often left unmentioned, it should be noted that in the scenarios described above, owing to the absence of cheap baseload power, electrolysis plants need to be able to bear dormancy (both technologically and economically) at certain times. In some cases, the capacity factor for electrolysis plants may only be between 40 and 60%. In general, gigawatt capacities installed on the production side need an aligned flexible demand. On a European scale, analyses show that 1600 gigawatt electrolysis and 7500 terawatt-hours of chemical storage may be needed to completely decarbonize heavy-duty transport such as trucks, ships, and planes (*2*). If this demand for energy storage were to be delivered from batteries, a capacity equivalent to that of ~50 billion Tesla Model 3 batteries would be needed (which is roughly 160 times the number of cars in Europe today). Storage costs for chemical energy as hydrogen, methane in caverns, or liquids are today at the level of <1 euro per kilowatt-hour (kWh) (excluding conversion costs), whereas the cost of battery storage is expected to remain considerably higher, reaching 80 euros/kWh by 2030 (*8*). Chemical liquid storage has much lower costs than battery storage but also lower cost than thermal and gaseous energy storage (*8*). Altogether, this analysis supports conversion of renewable electrical energy via electrolysis and storage of energy in liquid chemicals.

Electrolysis technologies will play a central role in future energy systems, acting as a vital link between the electric, gas, and thermal grids and providing fuel for the transportation sector. The huge scale at which electrolysis needs to be deployed means that there is undoubtedly room for a wide range of electrolysis technologies, notably alkaline, polymer electrolyte membrane (PEM), and solid oxide electrolysis cell (SOEC) technologies. Here, we argue that SOEC technology is the most suitable for wider-scale adoption. In the next sections, we outline the inherent advantages of SOECs, including the higher conversion efficiencies and raw material availability, as well as potential synergies from integration with downstream chemical syntheses. Furthermore, we present an overview of the advances in performance and durability on the cell and stack levels that we have witnessed over the past 10 to 15 years. In books written on electrolysis technologies, it is common practice to focus on alkaline and PEM technologies,

[1]Department of Energy Conversion and Storage, Technical University of Denmark, Fysikvej, Building 310, DK-2800 Kgs. Lyngby, Denmark. [2]Haldor Topsoe A/S, Haldor Topsøes Allé 1, DK-2800 Kgs. Lyngby, Denmark. [3]Energinet.dk, Tonne Kjærsvej 65, DK-7000 Fredericia, Denmark. [4]Department of Planning, The Technical Faculty of IT and Design, Aalborg University, A.C. Meyers Vænge 15.A, DK-2450 Copenhagen SV, Denmark.
***Corresponding author. Email: hauc@dtu.dk**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

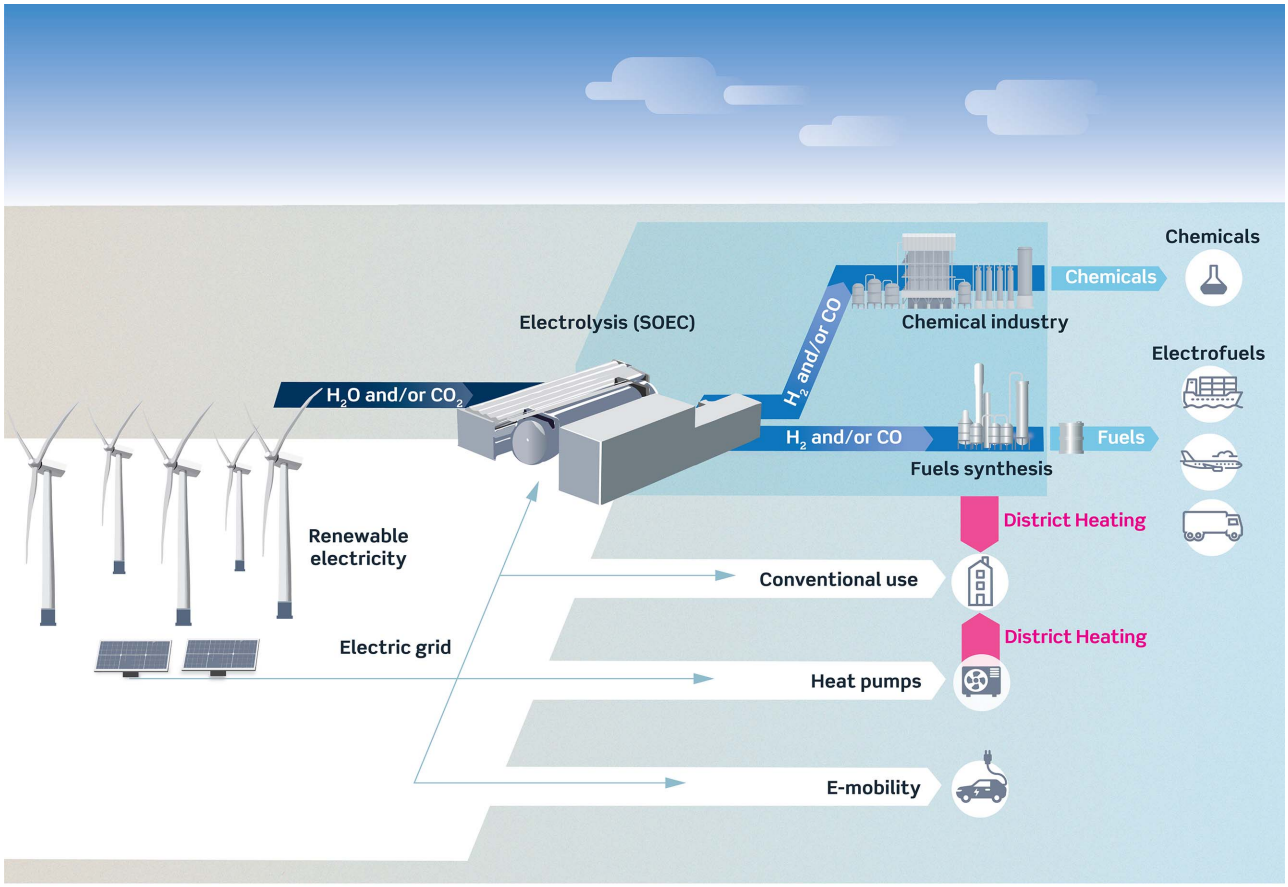


Fig. 1. Energy system based 100% on renewable energy. Illustration of the central role and integration of the SOEC technology in a future energy system with increased share of intermittent electricity from renewable sources such as wind and solar. The different branches of the energy system (electricity, gas, thermal grids, and transportation) are tightly coupled.

disregarding SOEC as an immature technology "still in R&D phase" (*9*). The aim of this Review is also to substantiate that, owing to recent improvements, SOECs are now ready for industrial scale-up and that this scale-up is in fact already happening at a rapid pace.

Unrivaled electrolysis efficiencies

An SOEC consists of three main components: two porous electrodes and a dense ceramic electrolyte capable of conducting oxide ions (O^{2-}). A scanning electron microscope image of a planar ceramic-based fuel electrode–supported cell is shown in Fig. 2A. SOECs are capable of splitting steam and CO_2 into H_2 and CO, respectively, but can also be operated in co-electrolysis mode, converting an H_2O-CO_2 mixture directly into synthesis gas (syngas; CO+H_2) (*10*). Electrochemical reduction of H_2O or CO_2 proceeds on the negatively charged fuel electrode, and oxide ions are conducted through the electrolyte and onto the positively charged oxygen electrode, where they oxidatively recombine into gas-phase O_2. When operated in reverse, an SOEC functions as a solid oxide fuel cell (SOFC). The cells may be operated reversibly, that is, switching between SOEC and SOFC modes, depending, for example, on the spot electricity price (*11*). SOECs are available in a wide range of structural designs (electrolyte-supported, metal-supported, ceramic fuel electrode–supported) and form factors (tubular, flat-tubular, planar), but the general characteristics of the technology remain the same regardless of the design. Multiple cells (~30 to 100) are combined in series in an SOEC stack by way of metallic interconnects, which provide electrical contact and gas separation between the cells as well as gas distribution across the cells. Stacks are built into modules and assembled in systems to reach the desired gas production rate (see below).

Elevated temperatures are necessary to reach sufficient ionic conductivity. The most common ceramic SOECs therefore typically operate at 600° to 850°C (*12*). The high operating temperature is an important feature of SOEC technology, giving rise to its two main advantages over competing alkaline and PEM electrolyzers, which typically operate at 60° to 160°C: more favorable thermodynamics and faster kinetics. Theoretical thermodynamic efficiency for both H_2O and CO_2 electrolysis increases with increasing temperature. As the temperature is raised from 25° to 800°C, the theoretical voltage for splitting of H_2O or CO_2 falls by 20 to 30% [see details in (*13*)]. In practice, when both thermodynamics and kinetics are considered, temperature-related efficiency gains are far higher (Fig. 2B): An SOEC operated at thermoneutral potential for the splitting of steam (1.29 V) will attain an electrolysis current density of ~1.5 A/cm^2, whereas a PEM electrolyzer operated at thermoneutral potential for the splitting of liquid water (1.47 V) attains a current density of ~0.5 A/cm^2 at similar gas compositions (*13*). PEM electrolyzers are typically operated at higher potentials (1.6 to 1.7 V), reaching current densities of ~1 A/cm^2 (*14*–*16*). The difference in thermoneutral potentials is directly related to the heat of evaporation, ΔH_{evap}. High-temperature operation is even more beneficial in CO_2 electrolysis mode, as evident from Fig. 2C. Lower cell voltage translates directly into lower operational costs (lower electricity demand per quantity of produced gas), while higher current densities are associated with lower capital costs (fewer electrolyzers needed to achieve the required capacity for gas production). Hence, the economic motivation

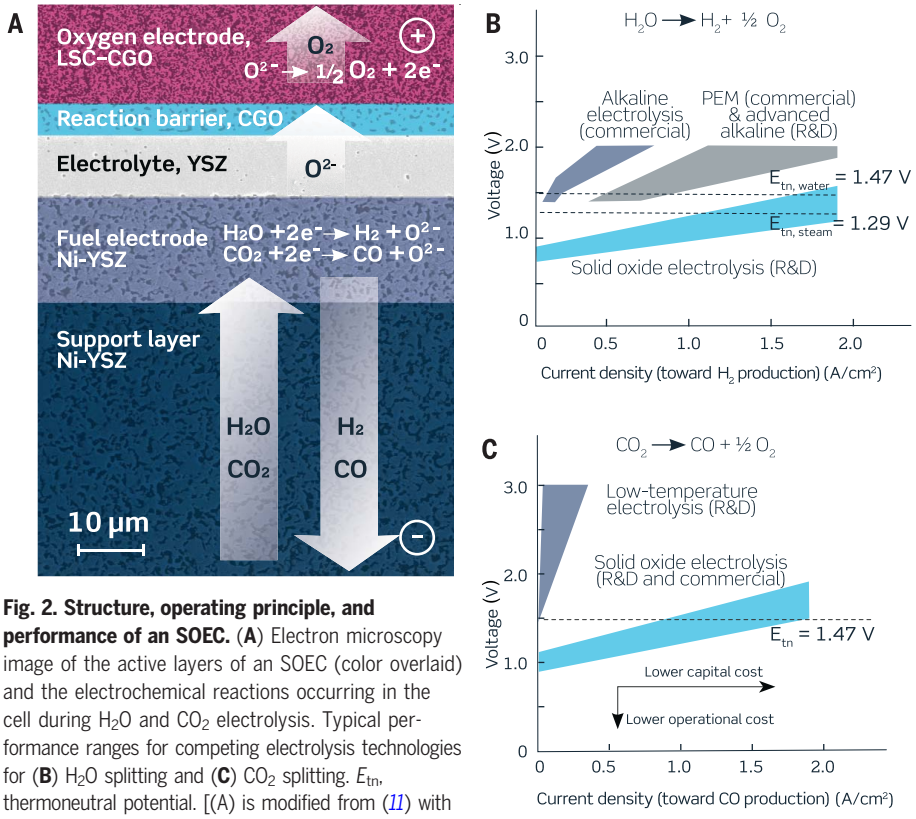


Fig. 2. Structure, operating principle, and performance of an SOEC. (**A**) Electron microscopy image of the active layers of an SOEC (color overlaid) and the electrochemical reactions occurring in the cell during H_2O and CO_2 electrolysis. Typical performance ranges for competing electrolysis technologies for (**B**) H_2O splitting and (**C**) CO_2 splitting. E_{tn}, thermoneutral potential. [(A) is modified from (*11*) with permission from Oxford University Press; (B) is adapted and modified from (*81*) with permission from Elsevier; and (C) is adapted from an open-access source (*54*)]

for wider adoption of SOEC technology remains high, provided that the main challenges (performance, degradation, and scale-up) are successfully addressed, as outlined in the subsequent paragraphs.

Technology based on abundant raw materials

The scale at which electrolysis cells will have to be deployed in future energy scenarios necessitates that any viable electrolysis technology be based on Earth-abundant materials, which is the case for the commonly used SOEC materials. In most cell designs, the main cell component is either yttria-stabilized zirconia (YSZ) (in the case of electrolyte-supported cells) or a composite of metallic Ni and YSZ (for fuel electrode–supported cells) (Fig. 2A). YSZ, a solid solution of a few mol % of yttria (Y_2O_3) in zirconia (ZrO_2), remains the electrolyte material of choice. Both yttria and zirconia are abundant materials: solid oxide cells providing 1 TW of power in fuel cell mode would require just 1 month's worth of global ZrO_2 production and 21 months' worth of Y_2O_3 (*17*). To put these numbers into perspective: 24 hours of 1 TW of power generated using Li-ion batteries would require Li corresponding to ~160 years of Li production in 2012, and 24 hours of 1 TW of power provided by a PEM fuel cell system would require 53 months' worth of global Pt production (*17*). A similar back-of-the-envelope calculation suggests that availability of Ir will be a concern. A single low-temperature 500-MW CO_2 electrolysis plant, as envisioned by De Luna *et al.* (*18*), would, at the IrO_2 loadings used in (*19*), require roughly two-thirds of the global annual production of Ir (*17*). The SOEC fuel electrode relies on catalysis by Ni, a non-noble metal with a flexible supply. For less-demanding applications, oxygen electrodes based on abundant Sr-doped $LaMnO_3$ (LSM) may be used, while higher-performing applications require electrodes based on mixed conductors, such as lanthanum-strontium-ferrite-cobaltite (LSCF) or lanthanum-strontium-cobaltite (LSC) (*20*). Thin (0.1- to 5-µm) layers of gadolinia-doped ceria (CGO) are commonly used to prevent reaction between oxygen electrode materials and YSZ (*21*).

Cell-level advances

Initial performance

Current-voltage curves recorded in steam electrolysis reveal that the initial performance of SOECs has increased by more than a factor of 2.5 over the past 15 years (Fig. 3A), as evidenced by a drop in area-specific resistance from 0.71 ohm•cm^2 to 0.27 ohm•cm^2 at 750°C (*22*, *23*). Although the exact same set of recent data are not readily available for CO_2 electrolysis, literature results suggest that cells optimized for H_2O electrolysis generally perform almost as well in CO_2 electrolysis (*24*). The performance enhancements shown in Fig. 3A have been achieved through three main modifications: (i) improved processing of cell layers, especially the fuel electrode; (ii) switching from LSM to mixed conducting oxides (LSCF, LSC) in oxygen electrode; and (iii) increasing the surface area of electrochemically active phases, for example, by infiltration, atomic layer deposition (ALD), or pulsed laser deposition (PLD).

Improved processing

Cell fabrication has witnessed multiple small breakthroughs during the past 15 years for both porous electrodes and the dense gas-tight electrolyte. Large shares of the performance increase illustrated in Fig. 3 are due to enhanced performance of the fuel electrode. Despite this, the largest individual contribution to the cell resistance still often originates from the charge transfer reaction at the triple phase boundaries (TPBs) in the fuel electrode. At the TPB, the electron-conducting and electrocatalytically active Ni, the gas-conducting pore, and the oxide-ion conducting YSZ are adjacent to each other. Advances in imaging techniques have made it possible to quantify the concentration and three-dimensional (3D) distribution of TPBs in real electrodes (*25*–*28*), enabling further improvements in electrode performance (*29*).

Methods for applying orders-of-magnitude thinner electrolytes in the submicrometer range (*21*) and the development of alternative ionic conductors (such as doped ceria or zirconia with previously unused or multiple dopants) demonstrate that research in electrolyte materials—a field that is ~100 years old—can still unlock improved performance (*12*, *30*, *31*). Improved processing methods have enhanced the mechanical properties (Weibull strength and fracture toughness) of SOECs (*32*, *33*), allowing the thickness of the Ni-YSZ support layers to be reduced from the typical 0.5- to 1-mm range (*34*) to 0.3 to 0.6 mm (*35*), resulting in major cost reductions on the cell level. Recently, an SOEC stack was flight-qualified for a mission to Mars (*36*), demonstrating that brittle ceramic cells, when properly encapsulated, can withstand very challenging mechanical load scenarios.

Mixed ionic electronic conducting oxides as oxygen electrode material

The second reason for performance enhancements is related to the introduction of mixed ionic and electronic conductors (MIEC) as electrocatalytic material for the oxygen electrode. In composites of pure ion conductors (e.g., YSZ) and pure electronic conductors (e.g., LSM), the electrochemistry is limited to the TPB sites, that is, to interfaces where the gas phase is simultaneously in contact with both solid phases. By replacing the electronic conductor with an MIEC material (e.g., LSCF), the reaction zone

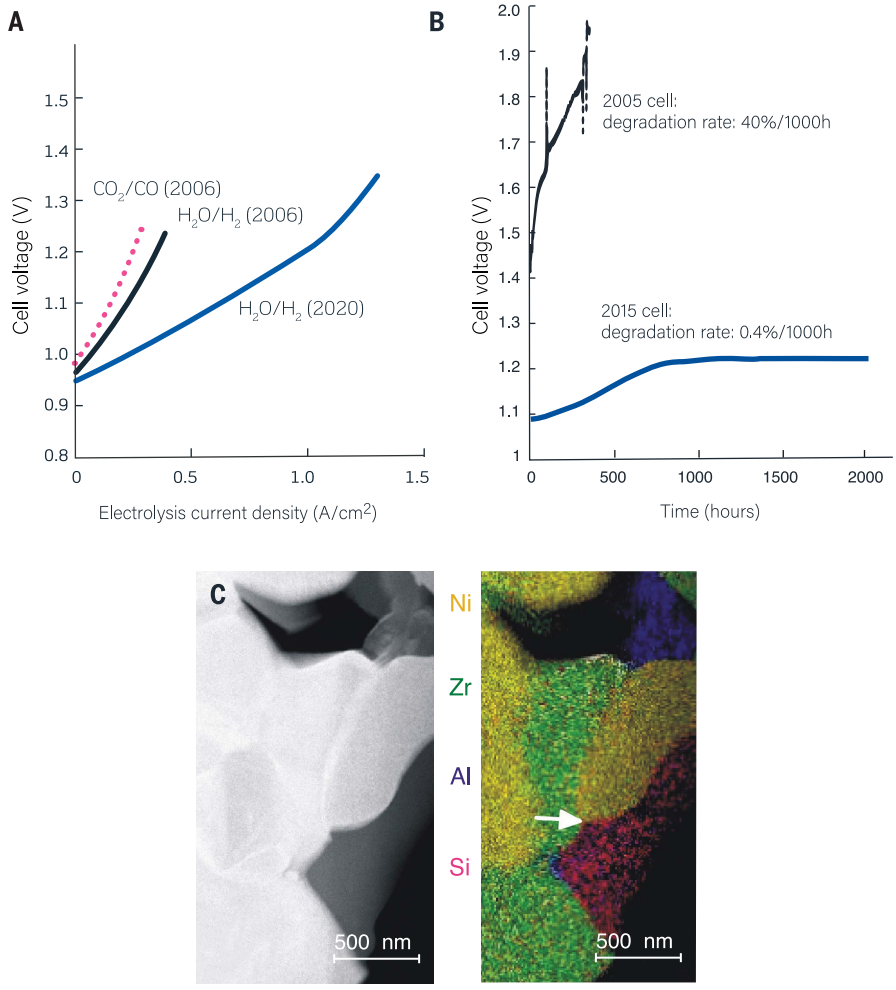


Fig. 3. Cell-level improvements. (**A**) Current-voltage curves for cells fabricated in 2006 and 2020 with data from (*22*, *23*, *82*) at 750°C, measured in H_2O/H_2 = 1 or CO_2/CO = 1. (**B**) Durability test of H_2O electrolysis at 1 A/cm^2 on a cell fabricated in 2005 measured at 850°C and a cell fabricated in 2015 measured at 800°C (*42*). All cells were supported by a Ni-YSZ electrode and had an active area of 16 cm^2. (**C**) TEM of Ni-YSZ electrode and corresponding energy dispersive x-ray mapping illustrating the presence of a Si-Al glass phase at the TPB (see white arrow) (*83*). [(C) is adapted with permission from the Electrochemical Society]

extends onto the MIEC phase, drastically increasing the number of electrochemically active sites in the oxygen electrode (*37*). MIEC generally also improves reaction kinetics.

Increased reaction zone via nanoengineering of electrode structures

A variety of methods can be applied to manufacture nanoengineered electrode structures, for example, infiltration, ALD, PLD, and exsolution, among others (*38*–*41*). Applying infiltrated nanosized electrocatalysts into a Ni-YSZ electrode structure shrinks the catalytic particles from micrometer scale to nanometer scale and drastically increases TPB lengths, thereby enhancing electrode performance. Compared with a standard electrode, electrodes with a larger number of active sites per electrode volume will inevitably be able to operate at lower overpotential for the same applied electrolysis current density. Furthermore, nanoparticle-infiltrated Ni-YSZ electrodes have been shown to be stable for hundreds of hours at high current densities (*38*).

Durability

Improvements in initial performance have been accompanied by tremendous advances in durability. Figure 3B shows two cell tests conducted at an electrolysis current of 1 A/cm^2. The resulting degradation rate, commonly expressed as voltage increase over 1000 hours, has decreased by a factor of 100 and is now below 0.5% per 1000 hours (*42*). Similar low degradation rates have recently been reported for cells operated at 1 A/cm^2 and lower temperatures (750°C) (*38*). Generally, cell degradation rates tend to decrease over time (Fig. 3B), a trend that has remained valid even in tests approaching 3 years in duration (*43*). From the increasing portfolio of SOEC long-term tests, two main insights with respect to degradation causes have emerged: (i) the effect of impurities and (ii) the close link between initial performance and resulting degradation rate.

It is the fuel electrode that is the primary source of degradation for most of the long-term tests reported. Several studies have shown that silica-containing impurities (glass phases) can block the electrocatalytically active sites, as illustrated in the transmission electron micrograph (TEM) in Fig. 3C. Such blocking of the TPB sites by nonconducting phases will inevitably cause degradation and lead to increased polarization resistance. Likewise, sulfur-containing impurities have been found to be critical in CO_2 and co-electrolysis experiments (*44*). For CO_2 and co-electrolysis, not only the thermodynamic threshold for carbon deposition but also the interplay between impurities (causing a higher fuel electrode overpotential) and the carbon deposition threshold play a role, as outlined in previous work (*45*); recent studies show positive effects from mixing of doped ceria and stabilized zirconia as oxide ion conducting material (*46*). The lesson learned from several long-term tests and impurity-related degradation causes is that cleaning of inlet gases is a beneficial lifetime-prolonging investment for SOECs, and gas purification can be, and is currently being, integrated in cell test as well as for demonstration plants (*44*).

Improved initial performance for the oxygen electrodes based on the MIEC electrocatalysts (e.g., LSC and LSCF) also positively affects the long-term durability. At higher current densities, lower-performing LSM-YSZ composite electrodes cause a higher pO_2 (partial pressure of oxygen) gradient in the YSZ electrolyte, prompting O_2 bubble formation in the electrolyte close to the oxygen electrode, which in turn leads to decreased O^{2-} conductivity and an irreversible increase in ohmic resistance of the cell (*47*). Although state-of-the-art LSCF- and LSC-based oxygen electrodes commonly contribute only ~5 to 10% of the total resistance of the cell, and limited degradation is observed (*20*, *23*, *29*), recent model electrode studies show that future oxygen electrode development efforts should be targeted toward designing and controlling both microstructure and especially surfaces (chemical compositions and valence states), as these greatly influence the electrochemical performance (*48*–*50*).

For the fuel electrode, there is also a link between initial performance and the resulting degradation rate (*29*). Several studies have shown that during long-term operation at high fuel electrode overpotentials (~300 mV), the percolating Ni network closest to the electrolyte is destroyed, and Ni migrates from the electrolyte-electrode interface to the support layer, resulting in irreversible loss of electrochemical performance (*29*, *51*). Similar studies of loss of the percolating Ni network were reported

in 3D reconstruction studies (52). Note that Ni migration is observed for cells operated in electrolysis mode but is not seen to the same extent in cells tested in fuel cell mode under similar test conditions, thus clearly pointing toward the local pO_2 in the fuel electrode as being the critical parameter (53). For future improvements of the SOECs, this observed irreversible degradation due to Ni migration needs to be tackled.

Stack-level improvements

Durability

A single 100 cm^2 solid oxide electrolysis cell operating at 0.8 A/cm^2 produces ~33 liters of gas (H_2 or CO) per hour. To increase the output capacity, electrolysis cells are connected in series and assembled together into stacks. Such stacks typically contain between 30 and 100 cells, translating into a production rate of 1 to 3 Nm^3 of gas (90 to 270 g H_2) per hour. Here, Nm^3 refers to normal cubic meters, i.e., cubic meters measured at 0°C and 1 atm. The electrochemical performance of SOEC stacks is determined not only by the activity of the cells but also by the performance of other stack components: metallic interconnects, glass sealings, and flow channels. The properties of each of these components change during long-term operation (interconnects corrode and creep, glass crystallizes, electrodes coarsen and accumulate impurities, and elements interdiffuse), and the components strongly interact. Thus, improvements demonstrated on a single-component level do not necessarily lead to enhanced performance on the stack level. However, substantial progress has been made in this regard, as evidenced by the gradually increasing test time reported in the literature (Fig. 4A) and by the corresponding drop in degradation rates (Fig. 4B). In 2011, the longest reported stack test in steam electrolysis lasted for <6 months, whereas now, stack lifetimes approaching 2.5 years have been experimentally demonstrated (43). It was only within the past ~5 years that tests exceeding 1000 hours were reported for CO_2 and co-electrolysis, but the durability gap between different electrolysis modes is closing fast, as more tests are performed with CO_2. Test times exceeding a year in dry CO_2 electrolysis mode suggest that SOEC is today by far the most mature technology for direct electrochemical conversion of CO_2 into CO (54).

State-of-the-art SOEC stacks are not only less prone to hard failures (sudden losses in performance) but, in line with cell-level results, the stacks also degrade less rapidly than a decade ago (Fig. 4B). For most reported tests, degradation rates now remain below 1% per 1000 hours, regardless of whether the stacks are operated in steam, CO_2, or in co-electrolysis mode. Encouragingly, stacks using similar cells and stack components have been operated continuously in fuel cell mode for more than 11 years (34), and >50,000 residential SOFC units have already been installed in Japan (55).

Robustness

A robust and flexible energy conversion technology is required to cope with the wide range of operating conditions in future energy scenarios. Historically, SOEC technology has struggled to demonstrate tolerance toward, for example, thermal (cool-down/heat-up) and emergency shutdown cycles, but this has recently changed. Stacks composed of ceramic-supported cells can now withstand ~150 thermal cycles (56), whereas stacks composed of metal-supported cells have been shown to endure >2500 such cycles (57). In fact, SOEC technology has matured to a point where the unreliability of system components (compressors, heaters, power supplies) other than the stack, or simply power failures, have become the main cause for system shutdowns (31, 58). Still, certain misconceptions prevail: first, that the brittleness of ceramic cells makes them unsuitable for pressurized operation, and second, that SOECs cannot follow dynamic load, for example, the intermittent electricity output from renewable energy sources. Both claims have been disproven by recent results, as outlined below.

Pressurized SOEC stack operation has been reported at pressures up to 25 bar (59, 60). Owing to improved kinetics and diffusion rates, the performance of SOECs typically improves slightly with increasing pressure. Stable operation at high pressures can successfully be maintained as long as the pressure difference between the fuel- and air-side compartments of the cell is minimized. High-pressure operation enables closer integration between electrolyzers and pressurized chemical synthesis reactors (31).

With respect to dynamic load operation, it has been shown that stacks at operating temperature can be ramped from zero to 80% load in a matter of milliseconds without damage (31). In a different study, an eight-cell stack was operated to follow a wind-profile for 1200 hours with the current load modified every 5 min without additional degradation (11). Similar tests have since then been performed using commercially available stacks (61), in co-electrolysis (62), and on both the single-cell level (63) and the system level (31). In large electrolysis plants like the one sketched in Fig. 1, stack-level load following may not be necessary; the modular structure of such plants allows

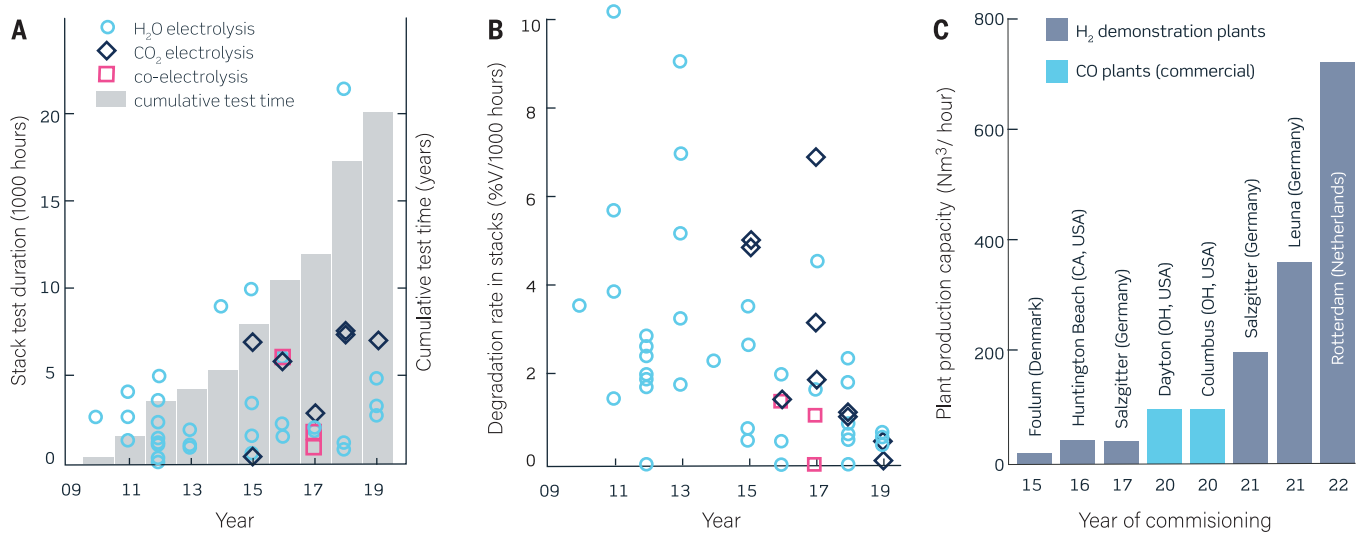


Fig. 4. Stack durability improvements and scale-up. Development over time of (**A**) reported stack test duration since 2009, (**B**) the corresponding degradation rates, and (**C**) SOEC plant production capacities from 2015 to 2022. Data in (A) and (B) are based on (84, 85), and data in (C) are based on (86).

the production rate to be adjusted incrementally by turning stack modules on or off.

The path to scale-up

The production capacity of an SOEC plant is directly proportional to the overall active area of the electrolysis stacks and to the applied current density. State-of-the-art SOEC stacks often have cell footprints (active electrode area) in the range of ~100 cm^2; however, cells with footprints as large as 550 cm^2 have been successfully produced and demonstrated (*35*). SOEC stacks with 100 or more cells are becoming commonplace, and designs for stacks with 350 cells have been proposed (*35*). Current densities as high as 3 A/cm^2 have been reported at the stack level using cells based on conventional SOEC materials (i.e., no noble metals), simply by optimizing the electrode microstructure and test conditions (*35*). A stack combining all of the listed improvements (550 cm^2, 3 A/cm^2, 350 cells per stack) would boast a production capacity of 240 Nm^3/hour of gas (21.6 kg H_2/hour), i.e., a factor of 80 to 240 higher than today's stacks. Although currently hypothetical, there are no fundamental obstacles preventing such stacks from being built.

Figure 4C illustrates the evolution of SOEC plant sizes measured in gas production capacity from 2015 to 2022, showing only plants for which funding has been allocated. As of early 2020, the largest operating SOEC demonstration plant is located in Salzgitter, Germany, and has a capacity of 40 Nm^3/hour of H_2. The plant consists of six modules of 48 stacks, each with 30 cells (*31*). However, demonstration plants with capacities reaching 720 Nm^3/hour H_2 (*64*) and multiple commercial (i.e., subsidy-free) plants with capacities of 96 Nm^3/hour CO (*65*) are to be commissioned in the next two years (Fig. 4C). Considering that the H_2 demand for a single methanol plant is on the order of 10,000 Nm^3/hour, it can be envisaged that SOEC systems can reach the required scale within a decade.

Outlook—from nanoscale to industrial scale

Synergies and integration of SOECs with chemical synthesis

From an overall efficiency point of view, electrolysis systems should be operated close to thermoneutral potential (1.29 and 1.47 V for steam and CO_2, respectively), that is, the potential at which the cooling effect from the endothermic electrolysis process is balanced by the Joule heating caused by the resistances in the cell [see (*13*)]. This in turn means that higher current density operation (higher H_2 and/or CO production rate) is possible for SOECs (Fig. 2, B and C), especially in the case of CO_2 electrolysis.

For the integration of SOECs in a future energy system as sketched in Fig. 1, substantial synergies can be obtained if the SOEC plant is integrated with downstream syntheses, for example, production of chemicals and fuels such as methane, methanol, gasoline, diesel, jet fuel, and ammonia (*66*, *67*). Figure 5 illustrates integration of SOECs with synthesis of methane, methanol, and ammonia and the beneficial use of heat from the synthesis processes, which are exothermic. The heat released can be used to produce the steam needed as feedstock for the SOEC. Note in Fig. 5C that for coupling with ammonia synthesis, the distinctive capability of the SOEC to function as an oxygen separation membrane is taken advantage of; the use of heat in lieu of power can be harnessed to eliminate the need for the expensive air separation unit to provide the nitrogen. Balance of plant details for the three scenarios sketched in Fig. 5 are given elsewhere (*68*). Recent demonstration projects with integration of SOEC systems with downstream synthesis processes are promising, for example, for upgrading of biogas to pipeline quality biomethane in a pilot plant providing 10 Nm^3/hour (*69*). Here, a constant production was maintained for >2000 hours with electric power consumption for the SOEC stack of 3.07 kWh/Nm^3 H_2, and the overall exergy efficiency from direct-current power to extra methane was close to 80% (*70*). Similarly, results on a 40 Nm^3/hour H_2 SOEC system as well as a smaller pressurized system to be integrated with a methanation unit have recently been reported (*31*).

Price competitiveness of SOEC systems

According to estimates given in literature, a 5-MW SOEC system is currently available at a cost (capital expenditure, or CAPEX) of roughly 2000 euros/kW, and this cost is projected to decrease to 1000 euros/kW by 2030 and 530 euros/kW by 2050, solely on the basis of cost reductions due to economies of scale (*71*). The operational cost (operational expenditure, or OPEX) will eventually be the main determining factor for choosing electrolysis technology. Studies have shown that electricity cost will be the major contributor (>70% in some cases) to OPEX in future electrified chemicals and synthetic fuel production plants based on electrolysis (*72*). However, at the end of the day, what the consumer will ask is "What is the price of my synthetic gasoline?" A recent technoeconomic assessment of power-to-gasoline conversion reached a gasoline price of 2.25 euros/liter with several cost and efficiency assumptions, including a price of electric power of 81.6 euros/megawatt-hour (*73*). Likewise, a targeted price of less than 2 euros/liter was announced for a demonstration project with a capacity of producing 8000 tons of liquid fuel per year (*74*).

Future R&D—on the cell, stack, and system levels

In this Review, we have highlighted some of the key developments in SOEC technology during

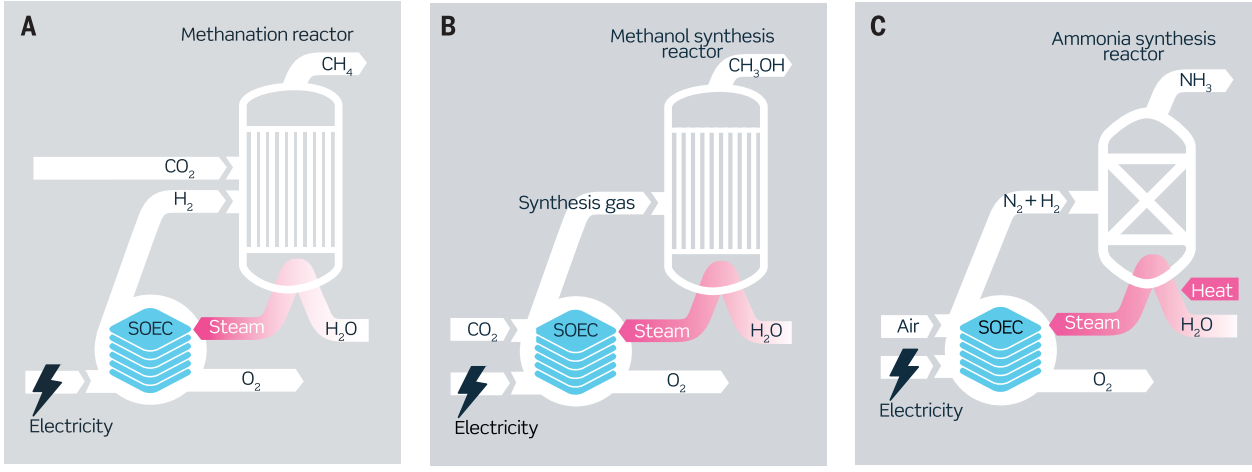


Fig. 5. Integration of solid oxide electrolysis systems with chemical syntheses. When SOECs are combined with a synthesis process, such as (**A**) methanation, (**B**) methanol synthesis, or (**C**) ammonia synthesis, reaction heat can be used to generate steam for SOECs. In (C), SOEC also functions as an oxygen-separation membrane, obviating the need for cryogenic air separation (*67*).

the past 10 to 15 years. The technology has moved from the research and development (R&D) phase into the demonstration and scale-up phase and is on the verge of commercialization. Nevertheless, R&D on SOECs has not stopped, and further performance enhancements, likely just as large as have been presented here, are possible on the cell, stack, and system levels.

On a cell level, over the past 15 years, the initial performance has increased by a factor of ~2.5, and the degradation rate has decreased by a factor of ~100. If there were only enough funding for R&D for a single part of the cell, it would be best spent on developing an even more robust and long-term stable fuel electrode. Projecting the next years of R&D on the fuel electrode, the core task will be to develop a robust electrode structure preventing Ni migration and, ideally, an impurity-tolerant electrode for which carbon deposition can be avoided. However, the issue of carbon deposition and impurity-induced degradation can to a large extent be avoided by purification of gas feedstock and proper choice of operating conditions on the system level, leaving Ni migration as the key issue to resolve. Work already initiated and referenced above on nanostructuring and examples of additions of dopant(s) to anchor Ni particles are promising (*75*). Furthermore, electrodes designed by exsolution in perovskite-structured titanates (e.g., La-Co-Ni-Ti oxides or La-Sr-Ti oxides with in situ exsolution of Mn, Cr, and/or Ni oxide nanoparticles) could be future viable routes to mitigate Ni migration and produce carbon-tolerant electrodes (*76*, *77*). Likewise, fuel electrodes based on nanoinfiltrated or exsolved nanoparticulate–doped Sr-Fe-Mo oxides could be alternative substitutes for the commonly used Ni-YSZ (*40*, *78*).

On the stack level, improving durability and reducing cost have served as the mantra for the solid oxide community ever since the first such electrolyzers were demonstrated in the 1980s (*79*). Going forward, the performance metric to be maximized should not be durability (expressed in hours) but rather the amount of gas produced over the stack's lifetime (*80*). This way, stacks operated at very different conditions (low or high current, dynamic or constant operation) could be compared. SOEC will not become the electrolysis technology of choice unless the total cost of ownership (cost of 1 kg H_2) is brought down to a lower level than that achievable by alkaline or PEM electrolyzers.

On the SOEC system and overall plant level, reliability of components other than the stack remains a challenge. As noted above, stack failures rarely initiate system shutdowns (*31*). Therefore, attention needs to be given to adapting and maturing balance-of-plant components for SOEC systems.

REFERENCES AND NOTES

1. B. V. Mathiesen *et al*., Smart Energy Systems for coherent 100% renewable energy and transport solutions. *Appl. Energy* **145**, 139–154 (2015). doi: 10.1016/j.apenergy.2015.01.075
2. D. Connolly, H. Lund, B. V. Mathiesen, Smart Energy Europe: The technical and economic impact of one potential 100% renewable energy scenario for the European Union. *Renew. Sustain. Energy Rev.* **60**, 1634–1653 (2016). doi: 10.1016/j.rser.2016.02.025
3. K. Hansen, B. V. Mathiesen, I. R. Skov, Full energy system transition towards 100% renewable energy in Germany in 2050. *Renew. Sustain. Energy Rev.* **102**, 1–13 (2019). doi: 10.1016/j.rser.2018.11.038
4. R. Perez, M. Perez, "A fundamental look at energy reserves for the planet" (2009); www.asrc.cestm.albany.edu/perez/Kit/pdf/a-fundamental-look-at%20the-planetary-energy-reserves.pdf.
5. B. V. Mathiesen, H. Lund, Comparative analyses of seven technologies to facilitate the integration of fluctuating renewable energy sources. *IET Renew. Power Gener.* **3**, 190–204 (2009). doi: 10.1049/iet-rpg:20080049
6. H. S. Edlefsen, A. B. Hansen, C. Vittrup, "System Perspective 2035. March 2018: Long-term perspectives for efficient use of renewable energy in the Danish energy system" (Energinet, 2018); https://en.energinet.dk/systemperspective2035.
7. B. V. Mathiesen *et al.*, "IDA's Energy Vision 2050: A Smart Energy System strategy for 100% renewable Denmark" (Department of Development and Planning, Aalborg University, 2015); https://vbn.aau.dk/ws/portalfiles/portal/222230514/Main_Report_IDAs_Energy_Vision_2050.pdf.
8. H. Lund, Energy storage and smart energy systems. *Int. J. Sustain. Energy Plan. Manag.* **11**, 3–14 (2016). doi: 10.5278/ijsepm.2016.11.2
9. A. Smith, A. S. Pedersen, "Electrolysis" in Encyclopedia of Environmental Management (Taylor and Francis, 2014).
10. C. Stoots, J. O'Brien, J. Hartvigsen, Results of recent high temperature coelectrolysis studies at the Idaho National Laboratory. *Int. J. Hydrogen Energy* **34**, 4208–4215 (2009). doi: 10.1016/j.ijhydene.2008.08.029
11. M. B. Mogensen *et al*., Reversible solid-oxide cells for clean and sustainable energy. *Clean Energy* **3**, 175–201 (2019). doi: 10.1093/ce/zkz023
12. E. D. Wachsman, K. T. Lee, Lowering the temperature of solid oxide fuel cells. *Science* **334**, 935–939 (2011). doi: 10.1126/science.1204090; pmid: 22096189
13. M. B. Mogensen, Thermodynamics of high temperature H_2O and CO_2 electrolysis, Figshare (2020); https://doi.org/10.6084/m9.figshare.12652322.v1.
14. S. Shiva Kumar, V. Himabindu, Hydrogen production by PEM water electrolysis – A review. *Mater. Sci. Energy Technol.* **2**, 442–454 (2019). doi: 10.1016/j.mset.2019.03.002
15. M. Carmo, D. L. Fritz, J. Mergel, D. Stolten, A comprehensive review on PEM water electrolysis. *Int. J. Hydrogen Energy* **38**, 4901–4934 (2013). doi: 10.1016/j.ijhydene.2013.01.151
16. S. Siracusano, V. Baglio, N. Van Dijk, L. Merlo, A. S. Aricò, Enhanced performance and durability of low catalyst loading PEM water electrolyser based on a short-side chain perfluorosulfonic ionomer. *Appl. Energy* **192**, 477–489 (2017). doi: 10.1016/j.apenergy.2016.09.011
17. P. C. K. Vesborg, T. F. Jaramillo, Addressing the terawatt challenge: Scalability in the supply of chemical elements for renewable energy. *RSC Advances* **2**, 7933 (2012). doi: 10.1039/c2ra20839c
18. P. De Luna *et al*., What would it take for renewably powered electrosynthesis to displace petrochemical processes? *Science* **364**, eaav3506 (2019). doi: 10.1126/science.aav3506; pmid: 31023896
19. S. Verma *et al*., Insights into the low overpotential electroreduction of CO_2 to CO on a supported gold catalyst in an alkaline flow electrolyzer. *ACS Energy Lett.* **3**, 193–198 (2018). doi: 10.1021/acsenergylett.7b01096
20. P. Hjalmarsson, X. Sun, Y.-L. Liu, M. Chen, Durability of high performance Ni–yttria stabilized zirconia supported solid oxide electrolysis cells at high current density. *J. Power Sources* **262**, 316–322 (2014). doi: 10.1016/j.jpowsour.2014.03.133
21. A. Ploner *et al*., Optimization of solid oxide cells and stacks for reversible operation. *ECS Trans.* **91**, 2517–2526 (2019). doi: 10.1149/09101.2517ecst
22. A. Hauch, S. H. Jensen, S. Ramousse, M. Mogensen, Performance and durability of solid oxide electrolysis cells. *J. Electrochem. Soc.* **153**, A1741 (2006). doi: 10.1149/1.2216562
23. X. Tong, S. Ovtar, K. Brodersen, P. V. Hendriksen, M. Chen, Large-area solid oxide cells with $La_{0.6}Sr_{0.4}CoO_{3-\delta}$ infiltrated oxygen electrodes for electricity generation and hydrogen production. *J. Power Sources* **451**, 227742 (2020). doi: 10.1016/j.jpowsour.2020.227742
24. S. Ebbesen, M. B. Mogensen, Electrolysis of carbon dioxide in solid oxide electrolysis cells. *J. Power Sources* **193**, 349–358 (2009). doi: 10.1016/j.jpowsour.2009.02.093
25. J. R. Wilson *et al*., Three-dimensional reconstruction of a solid-oxide fuel-cell anode. *Nat. Mater.* **5**, 541–544 (2006). doi: 10.1038/nmat1668; pmid: 16767095
26. S. Dierickx, J. Joos, A. Weber, E. Ivers-Tiffée, Advanced impedance modelling of Ni/8YSZ cermet anodes. *Electrochim. Acta* **265**, 736–750 (2018). doi: 10.1016/j.electacta.2017.12.029
27. Z. Jiao, N. Shikazono, Study on the effects of polarization on local morphological change of nickel at active three-phase-boundary using patterned nickel-film electrode in solid oxide fuel cell anode. *Acta Mater.* **135**, 124–131 (2017). doi: 10.1016/j.actamat.2017.05.051
28. G. Rinaldi *et al*., Post-test analysis on a solid oxide cell stack operated for 10,700 hours in steam electrolysis mode. *Fuel Cells* **17**, 541–549 (2017). doi: 10.1002/fuce.201600194
29. A. Hauch, K. Brodersen, M. Chen, M. B. Mogensen, Ni/YSZ electrodes structures optimized for increased electrolysis performance and durability. *Solid State Ion.* **293**, 27–36 (2016). doi: 10.1016/j.ssi.2016.06.003
30. P. Khajavi *et al*., Tetragonal phase stability maps of ceria-yttria co-doped zirconia: From powders to sintered ceramics. *Ceram. Int.* **46**, 9396–9405 (2020). doi: 10.1016/j.ceramint.2019.12.199
31. O. Posdziech, K. Schwarze, J. Brabandt, Efficient hydrogen production for industry and electricity storage via high-temperature electrolysis. *Int. J. Hydrogen Energy* **44**, 19089–19101 (2019). doi: 10.1016/j.ijhydene.2018.05.169
32. G. Pećanac, J. Wei, J. Malzbender, Fracture toughness of solid oxide fuel cell anode substrates determined by a double-torsion technique. *J. Power Sources* **327**, 629–637 (2016). doi: 10.1016/j.jpowsour.2016.07.103
33. D.-W. Ni *et al*., Influence of temperature and atmosphere on the strength and elastic modulus of solid oxide fuel cell anode supports. *J. Power Sources* **311**, 1–12 (2016). doi: 10.1016/j.jpowsour.2016.02.027
34. L. Blum *et al*., Long-term operation of solid oxide fuel cells and preliminary findings on accelerated testing. *Int. J. Hydrogen Energy* **45**, 8955–8964 (2020). doi: 10.1016/j.ijhydene.2020.01.074
35. T. Wood, H. He, T. Joia, in *Proceedings of 12th European SOFC & SOE Forum*, Lucerne, Switzerland, 5 to 8 July 2016, chap. 2, A1505, pp. 297–304.
36. J. Hartvigsen, S. Elangovan, J. Elwell, D. Larsen, Oxygen production from Mars atmosphere carbon dioxide using solid oxide electrolysis. *ECS Trans.* **78**, 2953–2963 (2017). doi: 10.1149/07801.2953ecst
37. S. B. Adler, Factors governing oxygen reduction in solid oxide fuel cell cathodes. *Chem. Rev.* **104**, 4791–4844 (2004). doi: 10.1021/cr020724o; pmid: 15669169
38. X. Tong, P. V. Hendriksen, A. Hauch, X. Sun, M. Chen, An up-scalable, infiltration-based approach for improving the durability of Ni/YSZ electrodes for solid oxide cells. *J. Electrochem. Soc.* **167**, 024519 (2020). doi: 10.1149/1945-7111/ab6f5c
39. Y. Cheng *et al*., Enhancing oxygen exchange activity by tailoring perovskite surfaces. *J. Phys. Chem. Lett.* **10**, 4082–4088 (2019). doi: 10.1021/acs.jpclett.9b01235; pmid: 31271532
40. H. Lv *et al*., In situ exsolved $FeNi_3$ nanoparticles on nickel doped $Sr_2Fe_{1.5}Mo_{0.5}O_{6-\delta}$ perovskite for efficient electrochemical CO_2 reduction reaction. *J. Mater. Chem. A.* **7**, 11967–11975 (2019). doi: 10.1039/C9TA03065D
41. J. T. S. Irvine *et al*., Evolution of the electrochemical interface in high-temperature fuel cells and electrolysers. *Nat. Energy* **1**, 15014 (2016). doi: 10.1038/nenergy.2015.14
42. A. Hauch *et al*., A decade of solid oxide electrolysis improvements at DTU Energy. *ECS Trans.* **75**, 3–14 (2017). doi: 10.1149/07542.0003ecst
43. J. Schefold, A. Brisse, H. Poepke, 23,000 h steam electrolysis with an electrolyte supported solid oxide cell. *Int. J. Hydrogen Energy* **42**, 13415–13426 (2017). doi: 10.1016/j.ijhydene.2017.01.072
44. S. D. Ebbesen, C. Graves, A. Hauch, S. H. Jensen, M. Mogensen, Poisoning of Solid Oxide Electrolysis Cells by Impurities. *J. Electrochem. Soc.* **157**, B1419 (2010). doi: 10.1149/1.3464804
45. T. L. Skafte, P. Blennow, J. Hjelm, C. Graves, Carbon deposition and sulfur poisoning during CO_2 electrolysis in nickel-based solid oxide cell electrodes. *J. Power Sources* **373**, 54–60 (2018). doi: 10.1016/j.jpowsour.2017.10.097
46. J. Wang *et al*., Threshold catalytic onset of carbon formation on CeO_2 during CO_2 electrolysis: Mechanism and inhibition.

J. Mater. Chem. A **7**, 15233–15243 (2019). doi: 10.1039/C9TA03265G
47. R. Knibbe, M. L. Traulsen, A. Hauch, S. D. Ebbesen, M. Mogensen, Solid oxide electrolysis cells: Degradation at high current densities. *J. Electrochem. Soc.* **157**, B1209–B1217 (2010). doi: 10.1149/1.3447752
48. Đ. Tripković, R. Küngas, M. B. Mogensen, P. V. Hendriksen, Surface recrystallization – an underestimated phenomenon affecting oxygen exchange activity. *J. Mater. Chem. A* **7**, 11782–11791 (2019). doi: 10.1039/C9TA02607J
49. J. Druce *et al*., Surface termination and subsurface restructuring of perovskite-based solid oxide electrode materials. *Energy Environ. Sci.* **7**, 3593–3599 (2014). doi: 10.1039/C4EE01497A
50. N. Tsvetkov, Q. Lu, L. Sun, E. J. Crumlin, B. Yildiz, Improved chemical and electrochemical stability of perovskite oxides with less reducible cations at the surface. *Nat. Mater.* **15**, 1010–1016 (2016). doi: 10.1038/nmat4659; pmid: 27295099
51. M. B. Mogensen *et al*., Relation between Ni particle shape change and Ni migration in Ni–YSZ electrodes – a hypothesis. *Fuel Cells* **17**, 434–441 (2017). doi: 10.1002/fuce.201600222
52. M. Trini *et al*., 3D microstructural characterization of Ni/YSZ electrodes exposed to 1 year of electrolysis testing. *J. Electrochem. Soc.* **166**, F158–F167 (2019). doi: 10.1149/2.1281902jes
53. M. Trini *et al*., Comparison of microstructural evolution of fuel electrodes in solid oxide fuel cells and electrolysis cells. *J. Power Sources* **450**, 227599 (2020). doi: 10.1016/j.jpowsour.2019.227599
54. R. Küngas, Review—Electrochemical CO_2 reduction for CO production: Comparison of low- and high-temperature electrolysis technologies. *J. Electrochem. Soc.* **167**, 044508 (2020). doi: 10.1149/1945-7111/ab7099
55. T. Nakao, S. Inoue, S. Uenoyama, Y. Takuwa, M. Suzuki, Progress of SOFC residential CHP system: Over 50,000 units market experience of Osaka Gas. *ECS Trans.* **91**, 43–49 (2019). doi: 10.1149/09101.0043ecst
56. N. P. Brandon, E. Ruiz-Trejo, P. Boldrin, Eds., *Solid Oxide Fuel Cell Lifetime and Reliability* (Academic Press, 2017).
57. P. Boldrin, N. P. Brandon, Progress and outlook for solid oxide fuel cells for transportation applications. *Nat. Catal.* **2**, 571–577 (2019). doi: 10.1038/s41929-019-0310-y
58. J. E. O'Brien *et al*., A 25 kW high temperature electrolysis facility for flexible hydrogen production and system integration studies. *Int. J. Hydrogen Energy* **45**, 15796–15804 (2020). doi: 10.1016/j.ijhydene.2020.04.074
59. M. Riedel, M. P. Heddrich, K. A. Friedrich, Experimental analysis of the co-electrolysis operation under pressurized conditions with a 10 layer SOC stack. *J. Electrochem. Soc.* **167**, 024504 (2020). doi: 10.1149/1945-7111/ab6820
60. S. H. Jensen, X. Sun, S. D. Ebbesen, M. Chen, Pressurized operation of a planar solid oxide cell stack. *Fuel Cells* **16**, 205–218 (2016). doi: 10.1002/fuce.201500180
61. M. Chen, Final Report, Project no. 12013, Solid Oxide Electrolysis for Grid Balancing (2020); https://energiforskning.dk/en/node/7317.
62. M. Rao, X. Sun, A. Hagen, Durability of solid oxide electrolysis stack under dynamic load cycling for syngas production. *J. Power Sources* **451**, 227781 (2020). doi: 10.1016/j.jpowsour.2020.227781
63. J. Schefold, A. Brisse, A. Surrey, C. Walter, 80,000 current on/off cycles in a one year long steam electrolysis test with a solid oxide cell. *Int. J. Hydrogen Energy* **45**, 5143–5154 (2020). doi: 10.1016/j.ijhydene.2019.05.124
64. Project MultiPLHY: Multimegawatt high-temperature electrolyser to generate green hydrogen for production of high-quality biofuels, Grant Agreement ID 875123, Horizon 2020, European Commission.
65. R. Küngas *et al*., Progress in SOEC development activities at Haldor Topsøe. *ECS Trans.* **91**, 215–223 (2019). doi: 10.1149/09101.0215ecst
66. J. B. Hansen, Solid oxide electrolysis—A key enabling technology for sustainable energy scenarios. *Faraday Discuss.* **182**, 9–48 (2015). doi: 10.1039/C5FD90071A; pmid: 26495443
67. J. B. Hansen, P. V. Hendriksen, The SOC4NH3 Project. Production and use of ammonia by solid oxide cells. *ECS Trans.* **91**, 2455–2465 (2019). doi: 10.1149/09101.2455ecst
68. J. B. Hansen, SOEC-based systems for Power-to-X, Figshare (2020); https://doi.org/10.6084/m9.figshare.12652373.
69. C. Dannesboe, J. B. Hansen, I. Johannsen, Catalytic methanation of CO_2 in biogas: Experimental results from a reactor at full scale. *React. Chem. Eng.* **5**, 183–189 (2020). doi: 10.1039/C9RE00351G
70. J. B. Hansen, F. Fock, H. H. Lindboe, Biogas upgrading: By steam electrolysis or co-electrolysis of biogas and steam. *ECS Trans.* **57**, 3089–3097 (2013). doi: 10.1149/05701.3089ecst
71. H. Böhm, A. Zauner, S. Goers, R. Tichler, P. Kroon, "Innovative large-scale energy storage technologies and Power-to-Gas concepts after optimization. D7.5: Report on experience curves and economies of scale" (2018); https://www.storeandgo.info/fileadmin/downloads/deliverables_2019/20190801-STOREandGO-D7.5-EIL-Report_on_experience_curves_and_economies_of_scale.pdf.
72. S. Siegemund, M. Trommler, P. Schmidt, "E-fuels study: The potential of electricity-based fuels for low-emission transport in the EU: An expertise by LBST and dena" (Article 9219, German Energy Agency, 2017); www.dena.de/fileadmin/dena/Dokumente/Pdf/9219_E-FUELS-STUDY_The_potential_of_electricity_based_fuels_for_low_emission_transport_in_the_EU.pdf.
73. L. Wang *et al*., Power-to-fuels via solid-oxide electrolyzer: Operating window and techno-economics. *Renew. Sustain. Energy Rev.* **110**, 174–187 (2019). doi: 10.1016/j.rser.2019.04.071
74. sunfire GMbH, "First commercial plant for the production of blue crude planned in Norway," press release (10 July 2017); www.sunfire.de/en/company/news/detail/first-commercial-plant-for-the-production-of-blue-crude-planned-in-norway?file=files/sunfire/images/content/company/press/2017_sunfire_PR_Nordic-Blue-Crude.pdf.
75. C. H. Law, S. W. Sofie, Anchoring of infiltrated nickel electrocatalyst by addition of aluminum titanate. *J. Electrochem. Soc.* **158**, B1137 (2011). doi: 10.1149/1.3610226
76. W. Wang *et al*., Enhanced carbon dioxide electrolysis at redox manipulated interfaces. *Nat. Commun.* **10**, 1550 (2019). doi: 10.1038/s41467-019-09568-1; pmid: 30948715
77. J. H. Myung, D. Neagu, D. N. Miller, J. T. S. Irvine, Switching on electrocatalytic activity in solid oxide cells. *Nature* **537**, 528–531 (2016). doi: 10.1038/nature19090; pmid: 27548878
78. Y. Wang, T. Liu, S. Fang, F. Chen, Syngas production on a symmetrical solid oxide H_2O/CO_2 co-electrolysis cell with $Sr_2Fe_{1.5}Mo_{0.5}O_6$-$Sm_{0.2}Ce_{0.8}O_{1.9}$ electrodes. *J. Power Sources* **305**, 240–248 (2016). doi: 10.1016/j.jpowsour.2015.11.097
79. W. Dönitz, E. Erdle, R. Schaumm, R. Streicher, Recent advances in the development of high temperature electrolysis technology in Germany. *Adv. Hydrog. Energy* **6**, 65–73 (1988).
80. R. Küngas *et al*., Lifetime capacity – an important performance metric for SOEC stacks. *ECS Trans.* **91**, 2601–2611 (2019). doi: 10.1149/09101.2601ecst
81. C. Graves, S. D. Ebbesen, M. Mogensen, K. S. Lackner, Sustainable hydrocarbon fuels by recycling CO_2 and H_2O with renewable or nuclear energy. *Renew. Sustain. Energy Rev.* **15**, 1–23 (2011). doi: 10.1016/j.rser.2010.07.014
82. S. D. Ebbesen, R. Knibbe, M. Mogensen, Co-electrolysis of steam and carbon dioxide in solid oxide cells. *J. Electrochem. Soc.* **159**, F482–F489 (2012). doi: 10.1149/2.076208jes
83. A. Hauch, J. R. Bowen, L. T. Kuhn, M. Mogensen, Nanoscale chemical analysis and imaging of solid oxide cells. *Electrochem. Solid-State Lett.* **11**, B38–B41 (2008). doi: 10.1149/1.2828845
84. T. L. Skafte, "Quantitative review of degradation and lifetime of solid oxide cells and stacks," *Proceedings of 12th European SOFC & SOE Forum*, Lucerne, Switzerland, 5 to 8 July 2016.
85. R. Küngas, Solid oxide electrolysis stack tests 2010-2019, Figshare (2020); https://doi.org/10.6084/m9.figshare.12137112.v1.
86. R. Küngas, SOEC plants 2015-2022, Figshare (2020); https://doi.org/10.6084/m9.figshare.12661994.v1.

ACKNOWLEDGMENTS

The authors acknowledge colleagues at the Department of Energy Conversion and Storage, the Technical University of Denmark, and Haldor Topsoe for valuable discussions. **Funding:** This work was partly funded by the Innovation Fund Denmark (IFD) under file no. 9067-00036B REFORGE. **Competing interests:** R.K. and P.B. own employee shares of Haldor Topsoe A/S. R.K., P.B., and J.B.H. are involved in the development and commercialization of SOEC-based electrolyzers and are (co-)inventors of several relevant patent applications assigned to Haldor Topsoe A/S.

10.1126/science.aba6118

RESEARCH ARTICLE SUMMARY

HIV

Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro

Devin E. Christensen*, Barbie K. Ganser-Pornillos*, Jarrod S. Johnson, Owen Pornillos†, Wesley I. Sundquist†

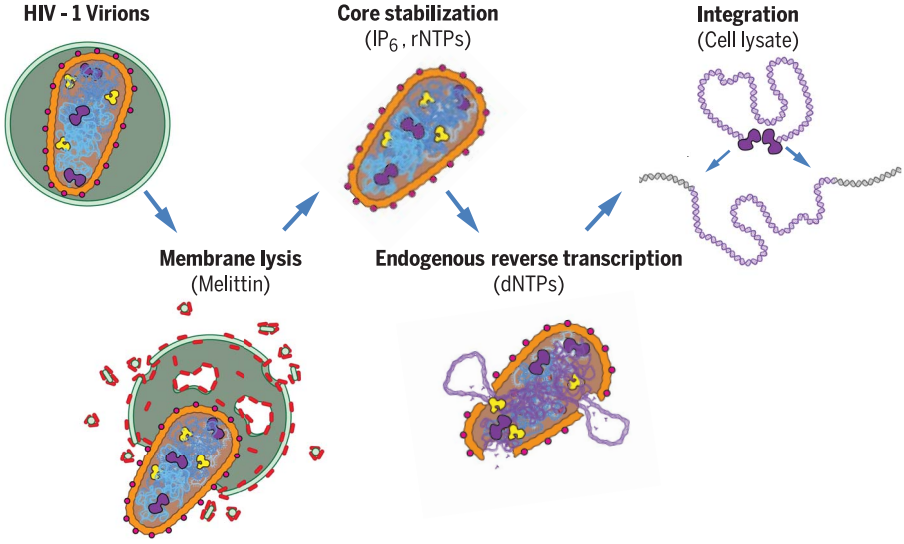
INTRODUCTION: Reverse transcription and integration are the signature events of retrovirus replication. Reverse transcription creates a double-stranded DNA (dsDNA) copy of the positive-sense viral RNA genome, and integration archives that copy within the genome of the infected cell. Both processes are targets of successful HIV-1 antiretroviral therapies, and the associated enzymatic activities have been characterized by elegant structural, biochemical, and molecular virological analyses. Nevertheless, mechanistic studies of these processes remain challenging because they are performed by viral core particles deep within the infected cell cytoplasm and nucleus. Of particular interest is defining whether the conical capsid that surrounds the viral RNA genome participates in the process of viral replication.

RATIONALE: In principle, informative mechanistic and imaging analyses of HIV-1 replication could be performed in vitro, but the coupled processes of reverse transcription and concerted integration have not yet been recapitulated outside of the cell. To address this limitation, we reconstituted these processes in a cell-free system, using purified HIV-1 virions as the source of viral genomes and enzymes.

RESULTS: We recapitulated the sequential processes of endogenous reverse transcription (ERT) and integration from viral core particles, which were released from purified HIV-1 virions by gently permeabilizing the viral membrane using a pore-forming peptide. ERT was initiated through addition of deoxynucleotide triphosphates (dNTPs), and the DNA products—early, intermediate, and late transcripts—appeared in high yields and in the expected temporal order, with late dsDNA products accumulating maximally after 8 to 10 hours. Integration of the resulting viral DNA into an exogenous target DNA was also recapitulated, provided that cell extract was added to the reaction mix. Deep sequencing and cloning confirmed that the 3′ and 5′ termini of the viral DNA were integrated in a concerted fashion, with the expected 5–base pair target site duplications in the target DNA, and with target site sequence preferences that resembled those reported for HIV-infected cells.

HIV-1 capsids are composed of hexagonal lattices of the viral CA protein, arrayed in a "fullerene cone" structure. Analyses of the requirements for efficient ERT and integration in our cell-free system revealed that these processes require the presence of capsid lattices of appropriate stability and geometry, as determined by using site-directed CA mutations, capsid-stabilizing factors, and potent new capsid inhibitors. Imaging with electron cryotomography revealed that many viral capsids remained largely intact during the ERT reaction. Capsid uncoating, when observed, did not occur in an all-or-none fashion, as might have been anticipated for a highly cooperative structure. Rather, uncoating proceeded through a continuum of disassembly intermediates in which portions of the capsid wall appeared lost in patches, as revealed through subunit-level lattice mapping. Largely intact capsids in which viral nucleic acid strands extruded through lattice openings were observed after 8 to 10 hours, which was coincident with the maximal accumulation of late ERT products and integration events.

CONCLUSIONS: We have reconstituted efficient reverse transcription and integration—the major early steps of the HIV-1 life cycle—in a cell-free system. Our data indicate that the viral capsid plays an active and indispensable role in supporting efficient reverse transcription. Thus, we consider the entire core particle, including the outer capsid shell, to be the true viral "replication complex." We further found that complete capsid uncoating may be a prerequisite for integration to occur. Thus, the capsid plays essential roles in the reactions that duplicate and archive the viral genome, in addition to previously established roles in protecting the viral genome from innate immune sensor surveillance and in helping the core to traverse the cytoplasm, enter the nucleus, and traffic to integration sites. We anticipate that our cell-free system will enable systematic analyses of the key steps in viral replication and integration and thereby elucidate the transformations that occur as HIV-1 proceeds through the first half of the viral life cycle. ■



Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. Images summarize stepwise reconstitution of highly efficient ERT and concerted integration in a cell-free system. Cryo–electron tomographic imaging showed that many core particles retained nearly complete outer viral capsids during the ERT reaction and that some lost surface patches, through which loops of the growing viral double-stranded cDNA could extrude. IP_6, inositol hexakisphosphate; rNTPs, ribonucleoside triphosphates.

The list of affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: opornillos@virginia.edu (O.P.); wes@biochem.utah.edu (W.I.S.)
Cite this article as D. E. Christensen *et al.*, *Science* **370**, eabc8420 (2020). DOI: 10.1126/science.abc8420

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abc8420

RESEARCH ARTICLE

HIV

Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro

Devin E. Christensen[1]*, Barbie K. Ganser-Pornillos[2]*, Jarrod S. Johnson[1], Owen Pornillos[2]†, Wesley I. Sundquist[1]†

During the first half of the viral life cycle, HIV-1 reverse transcribes its RNA genome and integrates the double-stranded DNA copy into a host cell chromosome. Despite progress in characterizing and inhibiting these processes, in situ mechanistic and structural studies remain challenging. This is because these operations are executed by individual viral preintegration complexes deep within cells. We therefore reconstituted and imaged the early stages of HIV-1 replication in a cell-free system. HIV-1 cores released from permeabilized virions supported efficient, capsid-dependent endogenous reverse transcription to produce double-stranded DNA genomes, which sometimes looped out from ruptured capsid walls. Concerted integration of both viral DNA ends into a target plasmid then proceeded in a cell extract–dependent reaction. This reconstituted system uncovers the role of the capsid in templating replication.

Reverse transcription and integration are the signature events of retrovirus replication. Reverse transcription creates a double-stranded DNA copy of the positive-sense viral RNA genome, and integration archives that copy within the genome of the infected cell, confounding viral eradication. Both processes are targets of successful HIV-1 antiretroviral therapies (*1*–*3*), and the associated enzymatic activities have been characterized by elegant structural, biochemical, and molecular virological analyses (*4*–*6*). Nevertheless, mechanistic studies of the native processes remain challenging because viral core particles perform their key transformations deep within the cell cytoplasm and nucleus. In principle, informative mechanistic and imaging analyses could be performed in vitro, but the coupled processes of reverse transcription and concerted integration have not yet been recapitulated outside of the cell. To address this limitation, we reconstituted these processes in a cell-free system, using purified HIV-1 virions as the source of viral genomes and enzymes.

Optimized endogenous reverse transcription

HIV-1 infections begin when the viral Envelope (Env) protein mediates fusion of the viral and target cell membranes, releasing the viral core particle into the cytoplasm. Core particles contain conical outer capsids comprising ~1500 CA protein subunits, arranged in a fullerene cone geometry composed of ~240 CA hexamers and 12 CA pentamers (*7*, *8*). The capsid encloses the factors required for replication, including the viral RNA genome (two copies per virion), NC protein (nucleocapsid, ~3000 copies), RT (reverse transcriptase, ~150 copies), and IN (integrase, ~150 copies) (*9*–*11*). Endogenous reverse transcription (ERT) of the viral RNA genome can initiate from a packaged host $tRNA^{Lys,3}$ primer when the viral membrane is permeabilized and deoxynucleotide triphosphates (dNTPs) are added (*12*). We used this system as a starting point for optimizing ERT efficiency, using $HIV\text{-}1_{SG3}$ virions that lacked the viral Env protein but were otherwise native and fully infectious when pseudotyped with viral envelope proteins (*13*–*15*). The viral membrane can be permeabilized without detergent by using the pore-forming melittin peptide from bee venom (*12*, *16*), and concentrations of melittin, buffer, and salt were optimized empirically (Materials and methods). Polyphosphorylated small molecules—including ribonucleoside triphosphates (rNTPs), dNTPs, and inositol hexakisphosphate (IP_6)—can bind and stabilize the hexameric CA building blocks of the viral capsid and increase ERT efficiency (*17*–*19*). We thus also included IP_6 at native cellular concentrations (40 μM) (*20*) and rNTPs and dNTPs at concentrations found in $CD4^+$ T cell cytoplasms (*21*, *22*). We also tested the effect of adding cell extracts, which have been reported to enhance ERT (*23*).

Successive stages of reverse transcription were monitored by means of quantitiative polymerase chain reaction (qPCR) by using primers that detected (i) an early initiation product that preceded minus strand strong stop DNA formation (MSSS) (Fig. 1A), (ii) an intermediate minus strand product that required first strand transfer (FST), or (iii) a late product that required second-strand transfer (SST) and initiation of double-stranded DNA (dsDNA) synthesis. At early time points, each input core could support formation of two single-stranded (MSSS) and intermediate (FST) transcripts because there are two strands of viral RNA per core. At later time points, qPCR measurements of these two transcripts could rise as high as four per input core because later replication steps create complementary DNA strands and also introduce second copies of the long terminal repeat (LTR) binding sites for the qPCR primers used to detect these transcripts (Fig. 1A). A double-stranded version of the qPCR primer binding site used to detect late (SST) products is only made once, however, and we presume that by this point in the replication cycle there will be fewer than two viral templates per core owing to strand switching (*24*). ERT efficiencies were therefore calculated by using qPCR to quantify late (SST) ERT products, assuming a theoretical yield of one late product per core and estimating input viral core numbers by quantifying virion CA levels and assuming an average of 3000 CA molecules per virion and 1.33 cores per virion (*9*, *11*, *25*).

ERT products increased with time and appeared in a temporal order that matched the well-characterized stages and timing of HIV replication in cells (*26*), with early reverse transcripts appearing rapidly (0.25 to 1 hour), intermediate FST products appearing after a short delay, and late reverse transcripts beginning to accumulate after a 3-hour lag and reaching a maximum after ~8 to 10 hours (Fig. 1B). ERT was reduced to background levels when the incubation was performed in the presence of the RT inhibitor efavirenz, or when HIV-1 virions contained RT enzymes with the inactivating D185A point mutation (Fig. 1C). As expected, ERT products did not form in reactions that lacked either melittin or dNTPs. IP_6 was also required, indicating a role for stable CA hexamers. Removing rNTPs reduced ERT efficiency modestly and had a more detrimental effect when IP_6 levels were lowered from 40 to 2 μM. Thus, rNTPs also contributed to ERT efficiency, particularly when IP_6 levels were limiting. Last, cell extract addition reproducibly increased ERT efficiency, from 0.39 ± 0.09 to 0.48 ± 0.08 late transcripts per core, implying that the extract contained a factor(s) that could stimulate ERT. The ~50% efficiency of this optimized ERT reaction was remarkably high given that production of late dsDNA transcripts requires multiple steps. The viral RT/ribonuclease H enzymes must copy sequences from both strands of the 9712–nucleotide (nt) genome, use two different RNA primers ($tRNA^{Lys,3}$ and the polypurine track), execute at least two different strand transfer steps, and degrade the genomic RNA template.

[1]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112, USA. [2]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22903, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: opornillos@virginia.edu (O.P.); wes@biochem.utah.edu (W.I.S.)

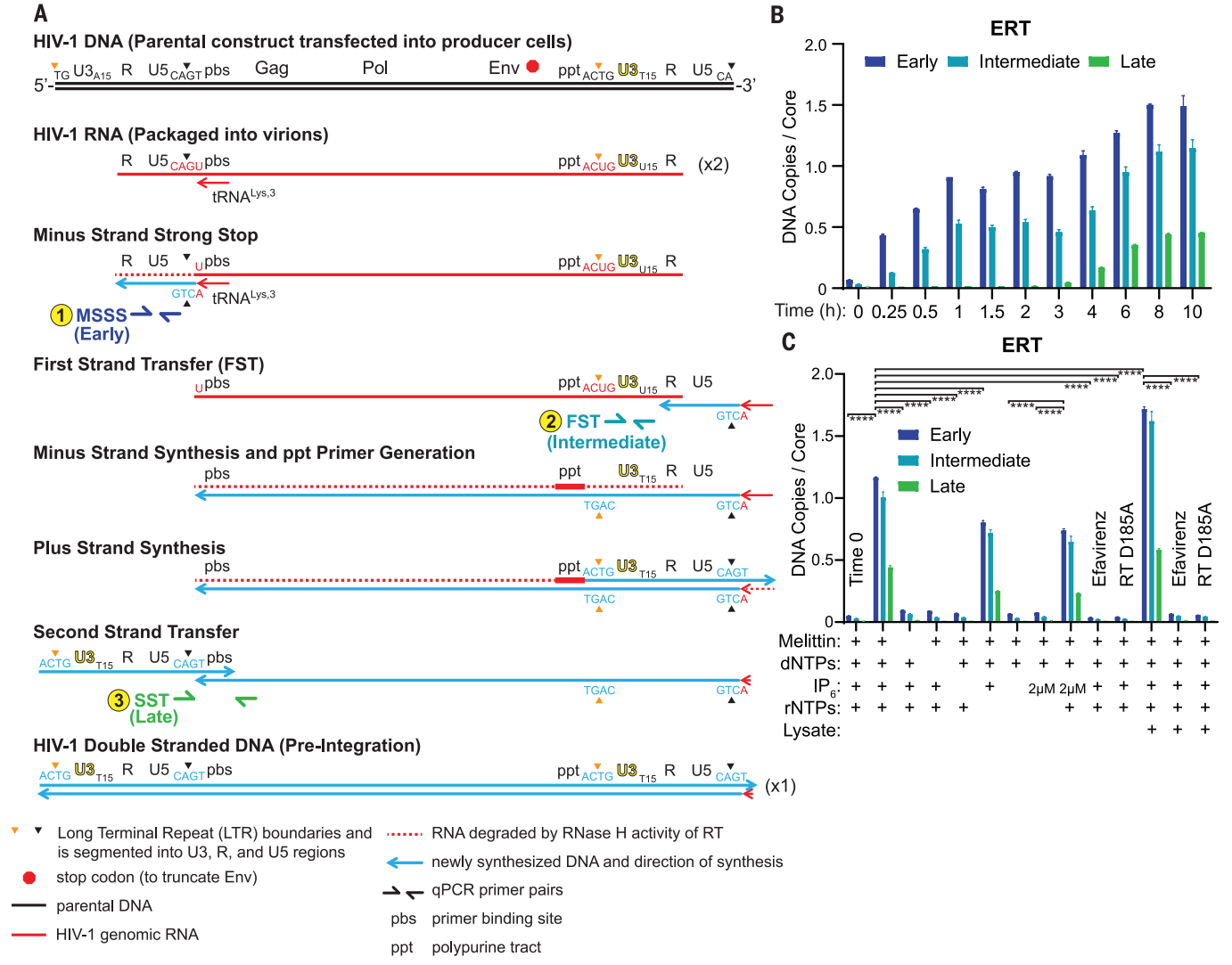


Fig. 1. Endogenous reverse transcription (ERT). (**A**) Different steps in HIV-1 reverse transcription and the primer positions used to assess them. (**B**) Time course showing accumulation of early, intermediate, and late ERT products, as quantified by means of qPCR. Synthesis of the viral plus strand cDNA introduces second copies of the binding sites for the primers used to detect early (MSSS) and intermediate (FST), which accounts for the apparent increase in signal for those two products seen between 4 and 8 hours. (**C**) ERT product accumulation under different reaction conditions (10-hours 37°C incubations except where noted) (Materials and methods). The RT inhibitor efavirenz (10 μM) and an inactive RT mutant enzyme (D185A) were used as negative controls. *P* values are from a one-way analysis of variance test with Tukey's multiple comparisons test: $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. Graphs and error bars show mean ± SD from three qPCR measurements from a representative experiment, selected from three independent experiments. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, D185A indicates that aspartic acid at position 185 was replaced by alanine.

Effects of viral capsid stability and capsid inhibitors

The HIV-1 capsid performs multiple functions, one of which may be to enhance the efficiency of reverse transcription (*27*, *28*). Consistent with such a role, CA mutations and small-molecule inhibitors that alter viral capsid stability can inhibit viral reverse transcription in infected cells (*29*, *30*). Similarly, the IP_6 and rNTP dependence of our ERT reactions suggested that the viral capsid might also be playing an important role in our system. To test this idea, we examined the effects of CA point mutations reported to either stabilize (E45A), moderately destabilize (Q63,67A), or severely destabilize the capsid (E28,29A) (*29*, *31*, *32*). The expected effects on capsid stability were confirmed by using negative-stain transmission electron microscopy (TEM) to quantify the numbers of wild-type and mutant cores with apparently intact capsids after 10-hour incubations under standard ERT conditions, but with three different IP_6 concentrations (Fig. 2A and fig. S1).

ERT efficiencies were also measured for each CA construct across a wide range of IP_6 concentrations (Fig. 2B). These experiments were performed in the absence of cell extract so that IP_6 concentrations could be controlled precisely. Wild-type capsids exhibited maximal ERT over an IP_6 concentration range of 40 to 80 μM, which approximates the estimated cellular concentration (~40 μM) (*20*). ERT efficiencies were reduced at both higher and lower IP_6 concentrations, ultimately reaching background levels in the absence of exogenous IP_6. The different CA mutants exhibited similar behaviors, but with differing optimal IP_6 concentrations that correlated inversely

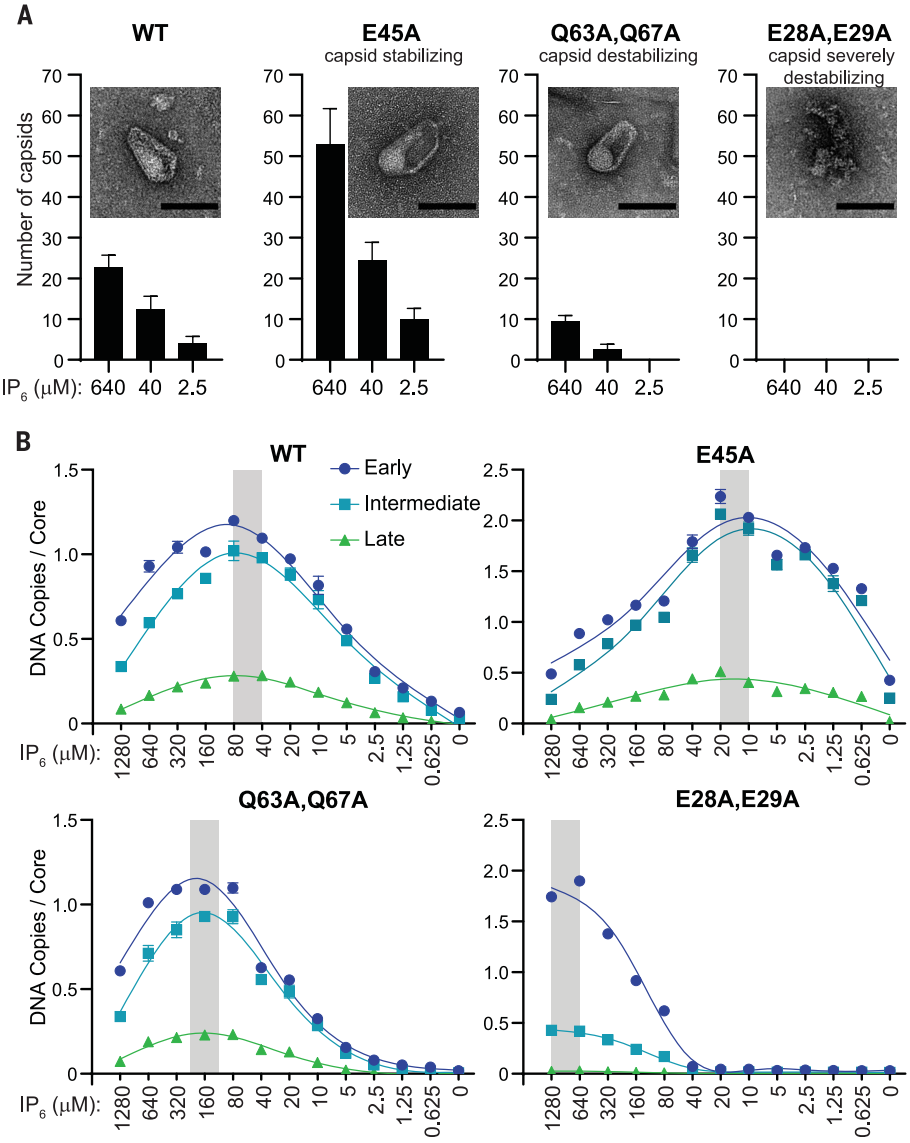


Fig. 2. Effects of capsid stability on ERT efficiency. (**A**) Stability of HIV-1 capsids composed of wild-type (WT) or the indicated CA mutant subunits. Capsid stability was assessed by treating virions with melittin to release cores into lysate-free ERT buffers that contained different IP_6 levels and then quantifying the numbers of apparently intact capsids by (insets) imaging with negative stain transmission electron microscopy (TEM). Scale bars, 100 nm. Graphs report the mean and standard deviation (n = 3) of intact cores per 900 μm^2 under the different conditions (Materials and methods) (fig. S1). (**B**) ERT levels from cores with WT or mutant CA proteins at different IP_6 concentrations. Gray bars highlight IP_6 concentrations that optimized transcript numbers. Standard "nonlysate" conditions were used for these experiments, except for variations in IP_6 levels.

with intrinsic capsid stability. IP_6 stabilizes capsids by binding within the central pores of the hexameric (and likely also pentameric) CA building blocks (*18*). Thus, IP_6-bound CA rings promote ERT, possibly by stabilizing higher-order capsid lattice assemblies. Conversely, the inhibition of ERT observed at high IP_6 levels, which was particularly evident for the hyperstable CA E45A mutant, indicates that capsid overstabilization is also detrimental, as has similarly been reported for HIV-1 infections in cultured primary $CD4^+$ T cells (*29*). Hence, a degree of capsid flexibility also appears to be required for efficient reverse transcription. The CA E28,29A mutant, which is severely impaired in forming intact capsids (*31*), initiated DNA synthesis efficiently when IP_6 levels were high, even under conditions in which no intact capsids were observed. Nevertheless, late transcripts were not produced under these conditions (Fig. 2B). Thus, local patches of capsid lattice, perhaps even just CA hexamers, appear to stimulate the production of early ERT products, and stable capsid lattices may be particularly important for efficient strand transfer or other later-stage steps in viral replication.

As another test of capsid involvement in ERT, we examined the effects of a potent small-molecule inhibitor that binds the viral capsid (*30*, *33*). GS-CA1 and its close analog GS-6702 bind in a pocket located between the N- and C-terminal domains of adjacent CA subunits within the hexameric ring and can thereby stabilize the CA hexamer. The inhibitors impair different capsid-dependent functions throughout different stages of the viral life cycle with distinct potencies, including virion production and maturation [GS-CA1 median inhibitory concentration (IC_{50}) 0.1 to 1 nM], reverse transcription (IC_{50} 5 to 25 nM), nuclear entry (IC_{50} 0.5 to 5 nM), and integration (IC_{50} 0.05 to 0.5 nM) (*30*, *33*). In our system, GS-CA1 strongly inhibited ERT in either the presence or absence of lysate (Fig. 3A), with an IC_{50} (10 to 100 nM) that matched that of GS-CA1 inhibition of viral reverse transcription in cells (*30*). A CA mutation (M66I) in the GS-CA1 binding site that confers drug resistance in cultured virus (*30*) also conferred GS-CA1 resistance in the ERT assay (Fig. 3B). Thus, GS-CA1 inhibition of viral reverse transcription and drug resistance were both recapitulated in vitro. The GS-CA1 inhibition patterns mimicked those seen for the severely capsid-destabilizing E28,E29A mutant at high IP_6 concentrations in that ERT initiated with reasonable efficiency even at high drug concentrations, whereas production of late transcripts was potently blocked. By contrast, the control inhibitor efavirenz, which acts most potently during RT initiation (*34*), blocked formation of early transcripts (Fig. 3, A and B).

Unlike the E28,29A capsid-destabilizing mutation, however, cores with recognizable capsids were readily observed with negative-stain TEM imaging of GS-CA1–treated wild-type samples. To determine how GS-CA1 treatment affected the integrity of the capsid structure at the subunit level, we used electron cryotomography (ECT) and subtomogram lattice mapping to visualize individual cores and map their constituent hexameric CA rings (fig. S2, A and B). All GS-CA1–treated capsids (18 of 18) had undergone substantial disassembly after 4-hour incubations (Fig. 3, C and D). By contrast, a majority of cores examined (8 of 14) in untreated control samples contained intact or nearly intact capsids. Thus, GS-CA1 appeared to accelerate loss of capsid integrity. Furthermore, comparison of surface curvature in the remaining capsid fragments indicated that GS-CA1 flattened the capsid lattice (fig. S2C). The GS-CA1 binding pocket (also called the "NTD-CTD interface") adopts multiple different configurations that help to generate capsid lattice curvature (*8*, *35*–*37*). GS-CA1 binding to the NTD-CTD interface may restrict the

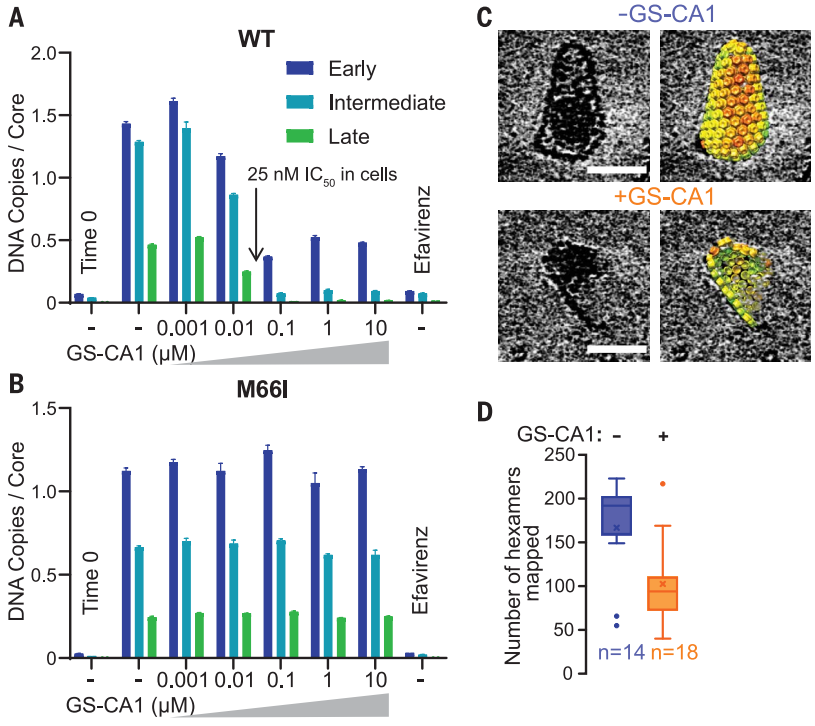


Fig. 3. Effects of the capsid inhibitor GS-CA1. (**A** and **B**) ERT levels in the presence of different concentrations of GS-CA1, for cores containing (A) WT or (B) a drug-resistant CA mutant M66I. Standard ERT conditions were used for these experiments. (**C**) Cryo-EM images of cores observed after 4 hours under standard ERT conditions in the absence or presence of 100 nM GS-CA1. Scale bars, 50 nm. The capsid lattices are rendered as hexamer units, each colored by cross-correlation value as determined with subtomogram averaging, with red indicating low correlations. Full galleries are shown in fig. S2. (**D**) Number of CA hexamers mapped per observable core in the absence or presence of 100 nM GS-CA1. Data are composite from two independent experiments that agreed well. Center lines show the medians, mean values are denoted by "×", box limits indicate the 25th and 75th percentiles, whiskers extend to minimum and maximum values, and circles indicate outliers, as determined in GraphPad Prism. Fully intact capsids contain an average of 240 hexamers.

flexibility of the CA subunits, which can explain the observed effect on lattice curvature. We propose that this effect in turn introduces lattice strain and promotes capsid fracturing. A similar capsid fracturing model has been proposed for PF74 (*38*), a less potent capsid inhibitor that also binds within the NTD-CTD interface.

Capsid uncoating

Many aspects of HIV-1 capsid uncoating are not well understood, including the timing, morphological transformations, and molecular trigger(s) [reviewed in (*28*, *39*–*42*)]. These uncertainties result, at least in part, from the challenges associated with TEM imaging of replicating viral cores within infected cells. To address this limitation, we used ECT and lattice mapping to image the viral capsids at the subunit level. Cores with mappable hexamers released from mellitin-treated virions in the absence of dNTPs and other reaction components (negative control) had predominantly intact or nearly intact capsids (45 of 55, 82%) (fig. S3A). By contrast, cores incubated under standard ERT conditions (37°C, 8 to 10 hours) exhibited a greater distribution of capsid morphologies (fig. S3, B and C). In these samples, cores surrounded by intact or nearly intact capsids still predominated among the cores with mappable capsid hexamers (53 of 94, 56%) (fig. S3B). We also observed a large number of partially disassembled capsid shells (34 of 94, 36%), with a continuum of lattice completeness that ranged from capsids with small ruptures to capsid remnants that were missing more than half of their hexamers (fig. S3B). The morphologies of these capsid disassembly intermediates were notable and unexpected in two ways. First, many capsids had lost local "patches" of their lattices, rather than disassembling in an all-or-none fashion (Fig. 4). Second, we could often discern strands of nucleic acids looping out of the resulting capsid openings (Fig. 4 and movies S1 to S3). It was possible to trace these continuous polynucleotide strands as they emanated from the core interior, exited through the gaps, and then looped back into the core. Loops were observed in the presence or absence of exogenous DNA, indicating that they were viral nucleic acids. We surmise that the looping strands corresponded to double-stranded cDNA transcripts, perhaps associated with NC protein, because the loops were only observed at late time points (8 to 10 hours), when most input cores had already completed reverse transcription. The longest loop shown in Fig. 4, top, would correspond to ~0.5 to 0.6 kb dsDNA, and the longest we observed in all of our images was ~1.5 kb, implying that the majority of the 9.7-kb genome remained tightly packed within the capsid. Viral capsid dissociation is promoted through reverse transcription (*26*, *43*), and our results support the proposal that increased pressure imposed by forming a relatively rigid dsDNA genome can rupture the viral capsid locally (*44*). Nevertheless, the capsid shell can remain largely intact during reverse transcription. This "local lattice rupture" model can reconcile reports that capsids become permeable to protein markers soon after entering the cytoplasm (*43*) yet remain largely intact all the way into the nucleus (*26*).

Integration

We also investigated whether our cell-free reactions could support integration of the reverse-transcribed viral dsDNA into a target DNA plasmid. Integration of the 3′ end of the viral genome was detected by using a PCR-based assay in which one primer annealed to the viral DNA and paired with one of two other primers that annealed to plasmid DNA sequences (figs. S4 and S5). This approach enriched for target plasmid integration events and detected integration in either possible orientation (*45*). In these experiments, the different stages of ERT again proceeded sequentially, with late dsDNA products accumulating maximally after 8 to 10 hours at 37°C (Fig. 5A, top). In the integration component of the assay, only low background qPCR signals were detected for the first 6 hours, before abruptly increasing >5000-fold at the 8- and 10-hour time points, indicating that fully reverse-transcribed viral dsDNA integrated into the target plasmid (Fig. 5A, bottom). In control experiments, integration signals were reduced to background levels in samples that contained efavirenz or an inactivating point mutation within IN (D116A) (Fig. 5B). As expected, efavirenz prevented the production of dsDNA integration substrates, whereas the IN D116A mutation did not alter ERT efficiency, indicating that this mutation specifically inhibited only the integration step. Integration also decreased in a dose-dependent fashion in the presence of increasing concentrations of the IN inhibitor raltegravir. ERT efficiencies were not altered by raltegravir treatment, indicating that the inhibitor also acted specifically at the integration step, and the IC_{50} (1 to

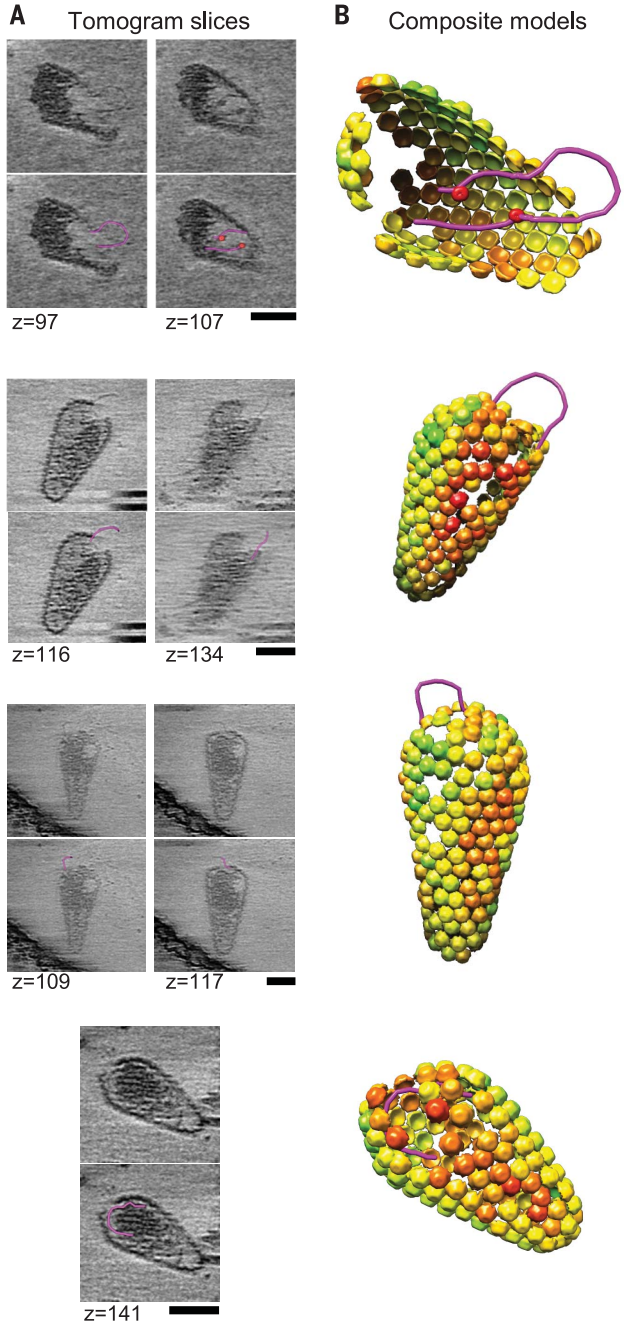


Fig. 4. Imaging cores during viral replication. (**A**) Tomographic slices of ruptured cores observed after 8 to 10 hours under standard ERT conditions. Polynucleotide loops (magenta) were segmented by use of the tools in imod (*58*). Red spheres indicate apparent protein densities associated with the loops within the core. Scale bars, 50 nm. (**B**) Composite representations illustrating the three-dimensional arrangements of loops and capsid lattice maps. The capsid lattice maps are rendered as hexamer units, each colored by cross-correlation value as determined with subtomogram averaging, with red indicating low correlations.

10 nM) value matched that seen in infected cells and in vitro assays (2 to 7 nM) (*46*, *47*). Integration also required cell extract (Fig. 5B), indicating that a host factor(s) was required to promote the reaction.

Integration efficiencies again exhibited strong biphasic dependence on IP_6 levels and capsid stability (fig. S6). Integration from wild-type capsids was strongly reduced at low levels of exogenous IP_6, where these capsids were unstable and late ERT product levels were very low, yet proceeded efficiently at low IP_6 levels when the capsids contained the stabilizing E45A CA mutant. Integration levels also diminished at very high IP_6 levels, where the wild-type capsids were hyperstabilized and late ERT product levels were again reduced. Integration was even strongly impaired at moderately elevated IP_6 levels where appreciable levels of late ERT products were present (for example, 320 μM IP_6), indicating that a final capsid "uncoating" reaction needs to occur before integration, as also occurs in the nucleus of HIV-1–infected cells (*26*).

Deep sequencing reactions confirmed authentic integration of the 3′ end of the viral dsDNA genome into the target plasmid (Fig. 5C, figs. S5 and S7, and table S1). Three independent reactions generated a total of 8.2×10^6 deep sequencing reads in which the 3′ end of the viral genome abutted plasmid DNA sequences. Collectively, these reads showed integration at nearly every base step of the entire 2423-nt target plasmid, in both possible orientations and with large read numbers at most base steps (mean 3397, median 307). Thus, many thousands of independent integration events occurred during each reaction. Integration frequencies varied dramatically at different sites, reflecting HIV-1 intasome sequence preferences (*48*–*50*), and the sequence preferences observed in vitro resembled those reported in HIV-1–infected cells (Fig. 5D) (*48*–*50*).

We also observed reads that corresponded to autointegration of the 3′ end of the genome back into the middle of the viral genome (fig. S8). These reads were much less abundant (1.1×10^4 reads), presumably because autointegration events occurred less frequently and were not exponentially amplified by our PCR protocol. Autointegration sites clustered about the center of the viral DNA and exhibited a strong orientation bias, which can be explained by the optimal DNA genome topology for concerted integration of both viral genome termini. Analogous off-pathway autointegration reactions also occur during viral infections, and their frequencies can be modulated by cellular host factors (*51*–*53*).

Creation of a viable provirus requires concerted integration of both viral termini across a 5–base pair (bp) staggered double-stranded break in the target DNA (fig. S4). Our cell-free system supported such concerted integration, as evaluated by amplifying, cloning, and sequencing individual plasmids that carried both integrated viral dsDNA termini (fig. S9). A selection of 99 independent clones from 12 different repetitions of this experiment were sequenced, and concerted integration of both viral termini was confirmed in every case. We also observed the expected nucleotide sequences at both ends of the viral DNA, with 5-bp target site duplications in the target DNA and with target site sequence preferences that again resembled those reported for HIV-infected cells (*48*–*50*). Moreover, a polymorphic sequence marker was transferred from the original 3′ viral RNA LTR element to the 5′ end of the dsDNA genome, which again is

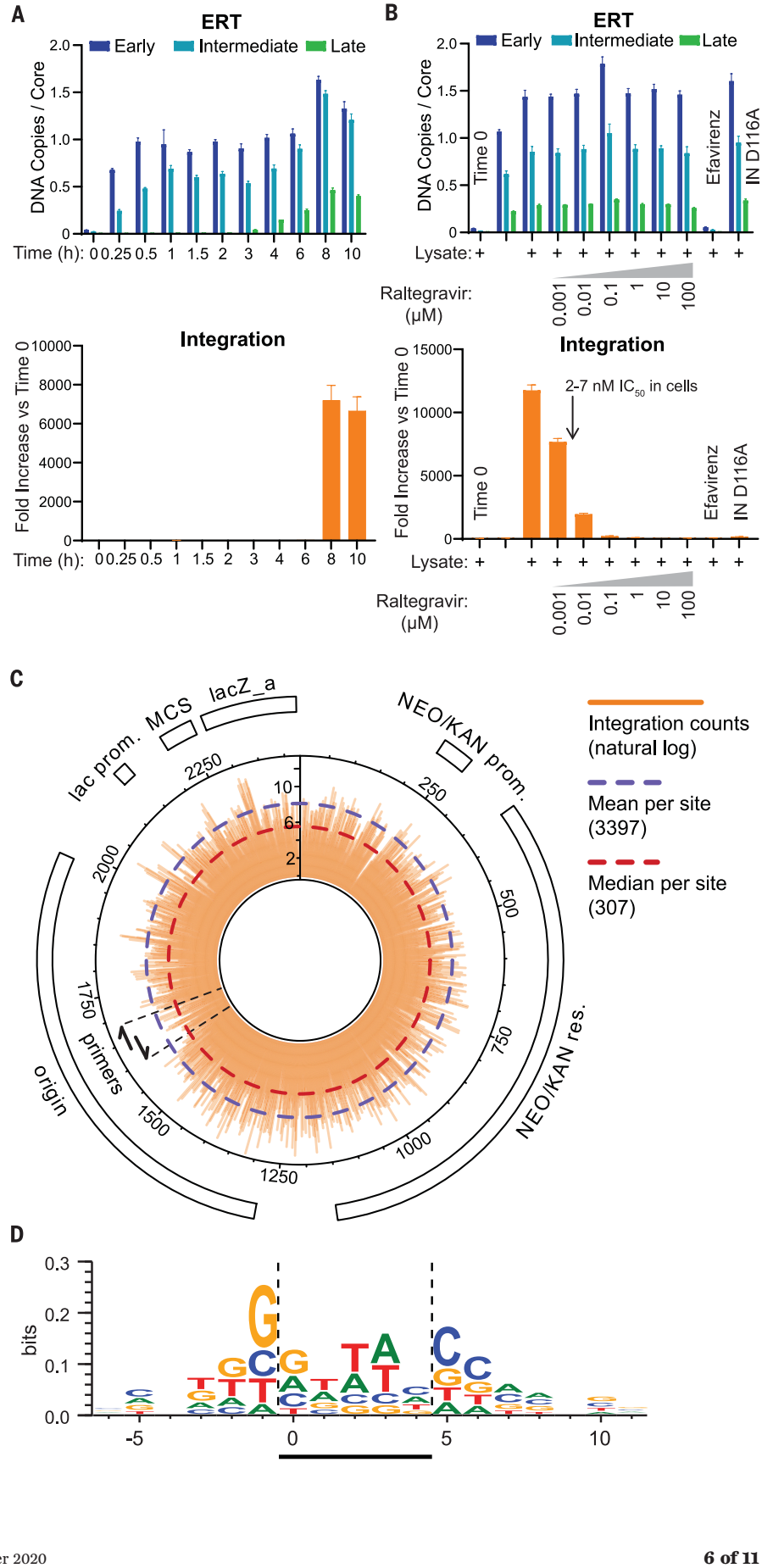


Fig. 5. Reconstituted integration. (**A**) Time courses showing (top) accumulation of ERT and (bottom) integration products. The time courses show the sequential formation of early, intermediate, and late ERT products, followed by integration. Integration products first appeared at time (*t*) = 8 hours, in parallel with peak production of late ERT products. Error bars here and in (B) show mean ± SD from three qPCR measurements from a representative experiment, selected from three independent experiments. (**B**) Quantification of (top) ERT and (bottom) integration products under different reaction conditions (Materials and methods). Negative controls show the effects of no incubation (*t* = 0), raltegravir (IN inhibitor, 1 nM to 100 μM), efavirenz (RT inhibitor, 10 μM), or an inactivating IN mutation (D116A). (**C**) Circular pK184 plasmid map showing the position and frequency of integration events determined through deep sequencing of read junctions between the 3′ end of the viral genome and the target plasmid (orange bars, ln scale). The graphic shows composite data from three independent integration and deep sequencing reactions, with integration in both possible orientations picked up by the two designated primers. (**D**) Favored DNA sequences for HIV integration, analyzed by use of WebLogo as described in Materials and methods. HIV integration in the target DNA sequence occurs between positions 0 and −1 on the sequenced strand and between positions 4 and 5 for the complementary strand. A representative logo from one of three experiments (table S1, sample 47) is shown. Bases and their heights indicate the level of conservation at each target base position (perfect conservation = 2 bits).

consistent with the established mechanisms for viral replication (Fig. 1A).

Discussion

We reconstituted the major early steps of the HIV-1 life cycle in a cell-free system. ERT was highly efficient, with approximately half of the input cores yielding late RT products, which required copying 10,000 nucleotides, degrading the viral RNA template, and executing two distinct sets of priming and strand transfer reactions. The virus has thus evolved mechanisms for performing each of the different replication steps very efficiently, apparently in response to the strong selective pressure imposed by productive infections being typically initiated by a single founder virus during sexual transmission (*54*). The viral CA protein is not an inert bystander during these processes but rather contributes directly to ERT efficiency. This activity requires intact CA rings and/or capsid lattice sections of the appropriate stability and geometry.

Our imaging studies revealed several interesting features of capsid-dependent ERT. Most fundamentally, we found that many viral capsids remained largely intact during the ERT reaction, and those that did uncoat did not do so in an all-or-none fashion, as might have been anticipated for a highly cooperative structure. Instead, we observed a continuum of disassembly intermediates. The most striking of these were largely intact capsids in which viral nucleic acid strands extruded through lattice openings. Thus, the initial stages of capsid uncoating may simply involve the loss of relatively small surface patches to accommodate the growing viral double-stranded cDNA, whose stiffness and persistence length will be greater than that of the templating single-stranded RNA genome. Such largely intact capsids could, in principle, explain the recent observation that reverse transcribing viral cores maintain the majority of their CA subunits as they move into the nucleus, complete reverse transcription, and traffic to sites of integration before uncoating fully (*26*). Local lattice ruptures might also help accommodate radial constriction of the capsid as the viral core passes through the narrow inner channel of the nuclear pore, which is currently a puzzling paradox in the field. Our experiments in soluble cell extracts demonstrate, however, that viral cores need not pass through nuclear pores to complete reverse transcription, uncoat, or integrate.

Although we have not yet systematically synchronized and imaged the different stages of reverse transcription, the extruded polynucleotide loops that we observed at late time points are consistent with a model in which much of the viral cDNA is not held in a strict orientation but where the viral genome termini reside in close proximity. This configuration could facilitate strand transfer reactions during replication and then position the termini for intasome formation and concerted integration. Association of the genome ends could occur when the RNA is first packaged, although at present, packaging is only known to require sequences located near the 5′ end of the viral RNA genome (*55*, *56*). Alternatively, the termini could associate during FST, when the MSSS cDNA copy of the 5′ end of the viral RNA genome reanneals to the 3′ end of the viral RNA genome. We envision that an extended capsid interior surface could "cage" RT to prevent dissociation (and promote reassociation), organize the genome termini, and/or facilitate key steps in viral replication such as priming, strand transfer, and intasome formation. Our experiments indicate that in order to function properly, this capsid surface must adopt the proper geometry yet retain some dynamic flexibility but also that the entire capsid need not remain fully closed. Last, the viral NC protein likely also facilitates replication and integration (*57*). Thus, layered contacts between the different core components appear to be sufficiently stable to ameliorate highly cooperative capsid uncoating and ensure that supramolecular viral complexes remain intact as they traffic through the cytoplasm and nucleus to sites of integration.

Our ERT reactions produced complete dsDNA reverse transcripts that were properly configured to access and use the packaged IN enzymes. Although we did not quantify integration efficiencies in absolute terms, thousands of integration events occurred in each reaction, and concerted integration was observed, indicating that native-like intasomes and synapses were formed (*4*). Integration was also dependent on an exogenous soluble cellular factor(s), whose identity will be of interest. Last, integration did not proceed efficiently when the capsid was hyperstablized, indicating that the CA lattice must rearrange or uncoat completely for integration to occur, as has been observed in the context of a viral infection (*26*). We anticipate that our cell-free system will enable systematic analyses of these and other key steps in viral replication and integration and thereby elucidate the transformations that occur as HIV-1 proceeds through the first half of the viral life cycle.

Materials and methods

Cell culture, virion production, and cell lysates

Human 293T and HeLa cells were obtained from the American Type Culture Collection and grown in Dulbecco modified Eagle medium (ThermoFisher) supplemented with 10% fetal calf serum (Atlanta Biologicals) at 37C in 5% CO_2. Cell cultures were confirmed to be negative for mycoplasma in routine tests (Lonza) every six months. Viral particles were produced by Calcium Phosphate transfection (Takara) of 293T cells with the pSG3$^{\Delta env}$ plasmid, obtained from the NIH AIDS Reagent Program (Catalog # 11051). Virions produced from this construct lack the viral Env protein, but are fully infectious when pseudotyped with envelope glycoproteins from HIV-1 or VSV [(*13*–*15*) and confirmed for our construct]. For virion production, 2.5×10^6 293T cells were seeded on 10 cm plates and transfected 24 hours later. 18 hours post transfection, media was removed and replaced with fresh media. 48 hours post transfection, virion-containing media was removed and filtered through a 0.45 μM membrane, layered on a 20% sucrose cushion in HS buffer (10 mM Hepes pH 7.4 and 140 mM NaCl), and centrifuged for 2 hours at 28,000 rpm in a SW32 Ti rotor (Beckman). After centrifugation, media was decanted and virion-containing pellets were resuspended in 100 μl HS buffer per 10 cm plate of transfected cells, aliquoted, flash frozen in liquid nitrogen, and stored at −80C. Virion CA concentrations were determined by HIV-1 p24 ELISA assay (XpressBio). Typical yields were 40 to 100 pmol CA per 10 cm plate.

HeLa cell lysates were prepared by seeding 2.8×10^6 cells on a 15 cm plate. After 48 hours, media was removed and cells were washed with PBS, removed from the plate with a cell lifter (Sigma Aldrich), placed in a microcentrifuge tube, and centrifuged for 1 min at 1500 rpm (4C). After centrifugation, excess PBS was removed and cells were resuspended in 200 μl of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 50 μg/ml melittin) per 15 cm plate and incubated on ice for 15 min. Lysates were then spun in a micro-centrifuge for 10 min at 17,000 rpm. The soluble fraction was collected, aliquoted, flash frozen in liquid nitrogen, and stored at −80C.

Lysates prepared in this fashion stimulated ERT (Fig. 1C) and were required for integration (Fig. 5B). Preliminary characterization of the extract activities indicates that flow throughs from lysates passed through Centricon filters with cutoffs of 3K, 50K, and 100K did not support the integration reaction, implying that the lysate is not simply supplying a small molecule like Mg^{2+} or IP_6. Integration activity was also lost when the lysates were heated to either 65C (20 min) or 95C (5 min), implying that at least one key factor is likely to be a structured biomolecule.

ERT assay

Prior to assaying for ERT, virion preparations were treated with cyanase (RiboSolutions) to digest any plasmid DNA carried-over from transfection. Briefly, virions were diluted to 20 nM CA in cyanase digestion buffer (50 mM Tris pH 8.0, 20 mM NaCl, and 12 mM $MgCl_2$) in a total volume of 800 μl. 1 μl cyanase was added and the reaction was incubated at 37C for 30 min. 60 μl of cyanase inactivation resin (RiboSolutions) was then added and incubated

with constant inversion at 37C for 20 min. Virions were then centrifuged for 1 min at 5000×g and virion containing supernatant was removed and stored on ice. Background level measurements in the ERT and integration assays were greatly reduced by cyanase treatment, demonstrating the importance of removing carry-over DNA from transfection.

ERT was initiated by combining 20 μl of virions with 20 μl of 2×ERT Buffer (50 mM Tris pH 7.4, 150 mM NaCl, 12 μM dCTP, 9.2 μM dGTP, 16 μM dTTP, 10.4 μM dATP, 50 μg/ml melittin, 80 μM IP_6, 360 μM CTP, 3.5 mM GTP, 1.4 mM UTP, 13.4 mM ATP, 6.67 mM $MgCl_2$, and 0.6 ng/μl of the pK184 target plasmid) and 5 μl of HeLa cell lysate or lysis buffer control, then incubated at 37C for 10 hours (denoted "Standard Reaction Conditions" or "Standard Reaction Conditions-Lysate", respectively), or otherwise for time courses as indicated in figures. rNTPs and dNTPs were purchased from Promega and their concentrations were chosen to correspond to those estimated for T lymphocyte cytoplasms (*21*, *22*). IP_6 was purchased from Sigma, and the optimal concentration was 40-80 μM, based upon yields of early (MSSS) DNA transcripts (Fig. 2A). 40 μM IP_6 was chosen for our standard condition because this is the estimated intracellular concentration (*20*). Melittin was purchased from Sigma, titrated, and used at a concentration that balanced the optimal yields of early (MSSS) transcripts, late (SST) transcripts, 3′ integration products and membrane-free cores. We found that Triton X-100 could work nearly as well as melittin in the ERT and integration assays, but the optimal concentration range was narrower and had greater variability in product yields. Mg^{2+} concentrations were also optimized, using early (MSSS) transcripts as the readout. Product formation was maximal in the range of 8-10 mM Mg^{2+}, and we chose a concentration (9.33 mM) that equalled the total rNTP concentration in the final ERT buffer conditions (i.e., rNTPs were added at 18.66 mM in 2× ERT buffer then diluted 1:1 with virions to 9.33 mM; $MgCl_2$ was added at 12 mM during Cyanase digestion and 6.67 mM in 2× ERT buffer, which when combined 1:1 resulted in a final concentration of 9.33 mM). 20 μl of 0.5 M EDTA was added to stop ERT. Final ERT DNA products were purified using a Qiaquick PCR purification kit (Qiagen) and eluted from the column in 120 μl elution buffer (Qiagen). Drugs at final concentrations were as follows; GS-CA1 (Gilead) in DMSO at 1 nM to 10 μM, raltegravir (NIH AIDS Reagent Program) in water at 1 nM to 100 μM, and efavirenz (NIH AIDS Reagent Program) in DMSO at 10 μM.

Two-step PCR 3′ integration assay

The plasmid pK184 (added to the 2×ERT Buffer, see above) was used as an integration target for the full-length viral cDNA produced in the ERT reaction. Purified reaction products (see above) were amplified by PCR using Phusion Polymerase (NEB), diluted 100-fold and measured by quantitative PCR (qPCR). Briefly, 5 μl of purified ERT product was combined with 10 μl Phusion HF Buffer, 200 μM each dNTP, 0.24 μM each of three primers (pK184-IN-FWD, pk184-IN-REV, and DC3014), 0.5 μl Phusion Polymerase, and water to a final volume of 50 μl. PCR cycles were as follows: Denaturation at 95C for 5 min; 23 cycles of 95C for 30 s, 58C for 30 s, and 72C for 90 s; and a final extension at 72C for 5 min followed by a hold at 4C. Final PCR products were purified with a Qiaquick PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in 120 μl elution buffer and diluted 100-fold for qPCR analysis of 3′ integration. Note that subsequent to these studies, re-optimization of the DC3014 primer further reduced the background PCR signal in the integration quantification assay by about 10-fold; the sequence of the improved primer (designated DC3014c) is provided in table S2.

To obtain 100 ng DNA for deep sequencing, 8 (for experimental repetitions 046 and 048) or 24 (for 047) identical PCR reactions were performed, combined, purified with a Qiaquick PCR purification kit (Qiagen), eluted in 50 μl elution buffer and submitted for deep sequencing analysis. Experimental repetitions 046 and 047 used 23 PCR amplification cycles, whereas 048 was amplified for 35 cycles. These samples were analyzed by deep sequencing to identify 3′ integration target site sequences.

Quantitative PCR analyses

Light Cycler 480 SYBR Green I Master (Roche) with a Light Cycler 480 (Roche) were used for all qPCR analyses. Briefly, 5 μl of either ERT product or diluted 3′ Integration PCR product, 500 nM each of two primers (MSSS-FWD + MSSS-REV, FST-FWD + FST-REV, Late RT-FWD + Late RT-REV, or AE3013 + IN990-REV for integration) and 10 μl of SYBR Green I Master Mix were combined in a total volume of 20 μl and analyzed by qPCR according to the manufacturer's instructions. All qPCR samples were analyzed in triplicate. Standard curves for ERT products were generated by using a 10-fold dilution series of the $pSG3^{\Delta env}$ plasmid for quantitation of ERT products. The number of DNA copies per core was determined from the ratio of viral cDNA copies (determined by qPCR) to the number of input core particles in each reaction (typically ~4×10^7). The number of HIV core particles in each reaction was determined from the input CA protein level (typically ~0.38 pmol, measured by ELISA, see above) and by assuming that there were 3000 Gag/CA molecules per virion and 1.33 cores per virion (*9*, *25*). Yields were calculated by assuming a theoretical yield of one late RT product/core. In reality, RT may initiate on both of the packaged RNA strands, and it is not clear when the conversion to a single template occurs. Note also that the primers used to quantify early/ MSSS and intermediate/FST products ultimately have two binding sites on the viral DNA template by the time the full dsDNA preintegration complex is formed. This effect is evident in the ERT time courses shown in Figs. 1A and 5A, where the qPCR products from these two primers rise late in the time course (e.g., compare the signals at 6 hours and 8 hours). These effects apparently explain why the qPCR measurements for the MSSS and FST primers can approach 2 copies/core in some experiments. Integration product levels were normalized relative to the qPCR signal observed for the t=0 time point. Three independent repetitions and analyses were performed, and representative examples are shown for the experiments in Figs. 1, Aand B; 2B; 3A;and 5, A and B; and fig. S7.

Negative stain transmission electron microscopy

ERT reaction mixtures (10 hours, 37°C, standard ERT conditions except that IP_6 was varied as described, 3.5 μl) were applied onto the carbon side of Formvar/Carbon-coated, 300-mesh copper grids (Electron Microscopy Sciences) for 90 s, rinsed for 90 s by flotation on a drop of 100 mM KCl, blotted dry, stained for 90 s by floating on a drop of 2% (w/v) solution of uranyl acetate, blotted dry, and allowed to air dry. Samples were viewed on an FEI Tecnai F20 transmission electron microscope operating at 120 kV. Core quantification was performed at a nominal magnification of ×11500, where viral capsids could be distinguished from similar size debris. For each grid, three well-stained squares were selected at random and capsids were counted manually while sweeping the perimeter of each square, covering an area of ~900 μm^2. Quantification was done in three independent experiments, and in two of these experiments the grids were randomized and the experimenter was blinded to their identities.

Cryo–electron tomography and lattice mapping

ERT reaction mixtures were mixed with an equal volume of 10-nm BSA Gold Tracer (Electron Microscopy Sciences), 3.5 μl aliquots were applied onto glow-discharged 2/2 carbon C-flat or Quantifoil grids (Electron Microscopy Science), blotted to near dryness manually, and then plunge-frozen into liquid ethane. Cryotomograms were acquired using an FEI Titan Krios electron microscope operating at 300 kV and equipped with either a Falcon III camera or a K3/GIF with a slit width of 20 eV. Tilt series were collected using the data collection software Tomography 4.0 (FEI) with an angular range of −60° to +60°, an angular increment of either 1 or 2°, defocus values of 5 to 10 μm,

and a nominal magnification of ×29,000 (Falcon III) or ×33,000 (K3), which correspond to a pixel (px) size of 2.92 (Falcon III) or 2.69 (K3) Å. Tilt series were aligned by using imod (*58*). Weighted back-projection or SIRT was used to reconstruct tomograms in imod.

Lattice mapping was performed based on previously described protocols (*59*, *60*) using as search template a cryoEM map of the HIV-1 CA hexamer (EMD-3465) (*8*) down sampled to 25 Å resolution. Each capsid surface was modeled and oversampled as a mesh by using surface modeling tools in the Dynamo software package (*58*). Two sets of sub-volumes centered at the same positions but of different box size (48 by 48 by 48 px^3 and 32 by 32 by 32 px^3) were extracted from 4× binned tomograms and assigned polar Euler angles in reference to the modeled capsid surface; azimuthal angles were randomized. For each box size set, four rounds of azimuthal angle search were performed, applying 6-fold symmetry, a low-pass filter of 30 Å and using a soft-edged cylindrical alignment mask in Dynamo. Overlapping and misaligned particles, as well as those with very low cross-correlation values were discarded. The two sets were then compared, and only particles common to both sets were retained. These were re-extracted (box size of 64 by 64 by 64 px^3) from 2× binned tomograms. To fill in gaps, the refined particle set was used to define a new surface model, which was oversampled and extracted (64 by 64 by 64 px^3) from 2× binned tomograms, retaining only those that are >10 px away from particles in the refined set. Four rounds of template alignment were then performed on the combined set of particles, using progressively narrower angular and positional search ranges and a low-pass filter set to 25 Å. After each round, overlapping and misaligned particles, as well as those with very low cross-correlation values were discarded. Final lattice maps were generated and visualized in Chimera (*61*) as described (*59*, *60*). Software used in this study were curated by SBGrid (*62*).

Deep sequencing and analysis of 3′ integration events

100 ng of the first PCR product from the 3′ integration assay (above) was submitted for Ion Torrent sequencing, performed at the DNA Sequencing Core Facility, University of Utah. Reads were screened for sequences with at least 20/23 matches to the extreme end of the HIV-1 3′ LTR using Seal (BBTools suite, Joint Genome Institute). To identify reads with junctions aligning to the pK184 target plasmid, these Seal reads were then remapped onto linearized pK184 reference genomes using BWA-MEM (*63*), again requiring at least 20/23 matches to the reference. To identify reads that could represent autointegration events, Seal reads with sequences matching the 3′ LTR were remapped onto a modified HIV-1 reference genome using BWA-MEM. Because the alignment software cannot distinguish identical sequences, the reference genome was modified by substituting the 5′ LTR with 633 Ns. Finally, subsets of reads from both types of alignments were checked manually to ensure that the alignment procedures identified integration sites and their orientations correctly. Alignments of integration events into pK184 and autointegration events into HIV-1 were visualized using the Integrative Genomics Viewer (*64*).

To score integration events at specific bases in the target plasmid, reads aligning to the HIV-1 3′ LTR and pK184 were parsed using their CIGAR string to identify the pK184 integration positions and orientations (Python code archived at GitHub repository DOI: 10.5281/zenodo.3966034). Because the alignment and integration analyses were performed on linearized pK184 reference genomes, the analyses were performed twice on different linearized pK184 reference genomes in which the plasmid "ends" were shifted to very different positions. After correcting for this shift, natural log-transformed merged integration counts were plotted using the circlize package in R (*65*). Autointegration events in HIV-1 were similarly scored by counting junction sites in reads that aligned to the 3′ LTR and to an additional site located between residues 1000 and 9000 of the HIV-1 genome. Autointegration counts were plotted using GraphPad Prism (*66*).

Cloning products of concerted integration

99 independent plasmids carrying concerted viral integration sites from 12 different integration reactions were cloned in two different ways: 53 clones were obtained following "repair" of the integration sites (described below), and 46 clones were obtained by simply amplifying the backbone without repair, implying that at least in some cases at least one of the integration sites must have been repaired in the extract. The repair protocol was as follows: 1) To remove the 5′ flap generated by integrase during integration, 10 μl of ERT product were incubated with 0.5 μl Thermostable FEN1 (NEB), 10 μl of Phusion HF Buffer, and 24.5 μl water at 65C for 5 min. 2) To fill in the 5 base integration target site gap, 200 μM each dNTP and 0.5 μl of Phusion Polymerase (NEB) was added, mixed, and incubated at 65C for 5 min. 3) To ligate the final phosphodiester bond and generate a complete double stranded DNA suitable for PCR amplification, T4 DNA ligase (NEB, 0.5 μl) was added and the sample was incubated at 37°C for 30 min.

Whether or not samples were repaired, plasmid backbones that carried the two viral termini were amplified by adding 240 nM of each primer (SG3-U3-XhoI-REV and SG3-U5-XhoI-FWD) to the sample to PCR amplify the two ends of the HIV genome along with the entire pK184 plasmid backbone (fig. S9). PCR cycles were as follows: Denaturation at 95C for 5 min; 30 cycles of 95C for 30 s, 58C for 30 s, and 72C for 2 min; and a final extension at 72C for 5 min followed by a hold at 4C. Final PCR products were purified with a Qiaquick PCR purification kit (Qiagen), digested with XhoI, ligated with T4 DNA Ligase (NEB) and transformed in *Escherichia coli* TOP10 cells. Plasmids were recovered from kanamycin-resistant colonies, and sequenced to verify concerted integration. In total, 151 plasmids recovered from 12 independent reactions were sequenced. Integration sites were judged to be necessarily independent if they did not match any of the other integration sites recovered from the same experiment (i.e., any duplicated integration sites from individual experiments were removed prior to analysis). Of the 103 remaining independent concerted integration events, all showed transfer of the 3′ $U3_{T15}$ to the 5′ LTR, 99 showed the expected viral termini and 5′ target site duplications, and four showed aberrant target site duplications of 22 bp (one case) and 65 bp (three cases, all at the same site).

Generation of integration consensus sites

To generate sequence logos showing target base preferences from our deep sequencing analysess of 3′ integration events, we first scaled integration counts to correct for PCR amplification bias using the gradients in read coverage around the pK184 target plasmid to guide the scaling. Read coverage depths, which decreased as a function of distance from the primer binding sites, were binned and plotted separately for forward- and reverse-oriented insertions into pK184. Forward and reverse coverage depth plots were used to generate LOWESS scaling curves and these curves were then applied to normalized integration count data from each independent experiment (using 50,000 as a normalization multiplier for forward and reverse integrations from samples 46 and 47, and 15,000 for forward and 25,000 for reverse integrations from sample 48 – these normalization multipliers were chosen to keep the total number of sequence strings analyzed for each sample below 1 million). Integration frequencies at each base step between 200 and 2200 base pairs away from each primer were then analyzed (this 2000 bp segment was chosen to exclude regions of low coverage immediately adjacent or far away from the primers). The scaled sequence strings were submitted to WebLogo 3 (*67*, *68*) to compute sequence logos. To generate the sequence logo for concerted integration, the sequences strings from the the 99 independent clones described above were submitted to WebLogo 3.

As shown in Fig. 5 and fig. S9, the consensus integration sequence motifs arising from both of our different sets of experiments differ slightly and are not perfectly palindromic, but are nevertheless clearly related to one another and to those reported previously (*48*–*50*). We have not interpreted our data further because our DNA target plasmid was limited in size and was not chromatinized, and because we do not know whether our extracts contain any, or all, of the host factors that might favor site selection in cells. We note, however, that our cell free system should provide a well-controlled platform for analyzing the effects of candidate host factors on integration site selection.

REFERENCES AND NOTES

1. M. S. Saag *et al*., Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2018 Recommendations of the International Antiviral Society-USA Panel. *JAMA* **320**, 379–396 (2018). doi: 10.1001/jama.2018.8431; pmid: 30043070
2. E. De Clercq, The nucleoside reverse transcriptase inhibitors, nonnucleoside reverse transcriptase inhibitors, and protease inhibitors in the treatment of HIV infections (AIDS). *Adv. Pharmacol.* **67**, 317–358 (2013). doi: 10.1016/B978-0-12-405880-4.00009-3; pmid: 23886005
3. E. J. Arts, D. J. Hazuda, HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161 (2012). doi: 10.1101/cshperspect.a007161; pmid: 22474613
4. P. Lesbats, A. N. Engelman, P. Cherepanov, Retroviral DNA integration. *Chem. Rev.* **116**, 12730–12757 (2016). doi: 10.1021/acs.chemrev.6b00125; pmid: 27198982
5. J. M. Coffin, H. Fan, The discovery of reverse transcriptase. *Annu. Rev. Virol.* **3**, 29–51 (2016). doi: 10.1146/annurev-virology-110615-035556; pmid: 27482900
6. S. G. Sarafianos *et al*., Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385**, 693–713 (2009). doi: 10.1016/j.jmb.2008.10.071; pmid: 19022262
7. O. Pornillos, B. K. Ganser-Pornillos, Maturation of retroviruses. *Curr. Opin. Virol.* **36**, 47–55 (2019). doi: 10.1016/j.coviro.2019.05.004; pmid: 31185449
8. S. Mattei, B. Glass, W. J. Hagen, H. G. Kräusslich, J. A. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434–1437 (2016). doi: 10.1126/science.aah4972; pmid: 27980210
9. J. A. Briggs *et al*., The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11**, 672–675 (2004). doi: 10.1038/nsmb785; pmid: 15208690
10. C. M. Swanson, M. H. Malim, SnapShot: HIV-1 proteins. *Cell* **133**, 742–742.e1, 742.e1 (2008). doi: 10.1016/j.cell.2008.05.005; pmid: 18485880
11. L. A. Carlson *et al*., Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**, 592–599 (2008). doi: 10.1016/j.chom.2008.10.013; pmid: 19064259
12. W. H. Yong, S. Wyman, J. A. Levy, Optimal conditions for synthesizing complementary DNA in the HIV-1 endogenous reverse transcriptase reaction. *AIDS* **4**, 199–206 (1990). doi: 10.1097/00002030-199003000-00004; pmid: 1693515
13. S. K. Ghosh *et al*., A molecular clone of HIV-1 tropic and cytopathic for human and chimpanzee lymphocytes. *Virology* **194**, 858–864 (1993). doi: 10.1006/viro.1993.1331; pmid: 8503191
14. X. Wei *et al*., Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **46**, 1896–1905 (2002). doi: 10.1128/AAC.46.6.1896-1905.2002; pmid: 12019106
15. X. Wei *et al*., Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003). doi: 10.1038/nature01470; pmid: 12646921
16. L. R. Boone, A. Skalka, Two species of full-length cDNA are synthesized in high yield by melittin-treated avian retrovirus particles. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 847–851 (1980). doi: 10.1073/pnas.77.2.847; pmid: 6153806
17. D. L. Mallery *et al*., IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife* **7**, e35335 (2018). doi: 10.7554/eLife.35335; pmid: 29848441
18. R. A. Dick *et al*., Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509–512 (2018). doi: 10.1038/s41586-018-0396-4; pmid: 30069050
19. D. A. Jacques *et al*., HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* **536**, 349–353 (2016). doi: 10.1038/nature19098; pmid: 27509857
20. A. J. Letcher, M. J. Schell, R. F. Irvine, Do mammals make all their own inositol hexakisphosphate? *Biochem. J.* **416**, 263–270 (2008). doi: 10.1042/BJ20081417; pmid: 18684107
21. T. L. Diamond *et al*., Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J. Biol. Chem.* **279**, 51545–51553 (2004). doi: 10.1074/jbc.M408573200; pmid: 15452123
22. E. M. Kennedy *et al*., Ribonucleoside triphosphates as substrate of human immunodeficiency virus type 1 reverse transcriptase in human macrophages. *J. Biol. Chem.* **285**, 39380–39391 (2010). doi: 10.1074/jbc.M110.178582; pmid: 20924117
23. D. Warrilow *et al*., Cell factors stimulate human immunodeficiency virus type 1 reverse transcription in vitro. *J. Virol.* **82**, 1425–1437 (2008). doi: 10.1128/JVI.01808-07; pmid: 18045931
24. J. M. O. Rawson, O. A. Nikolaitchik, B. F. Keele, V. K. Pathak, W. S. Hu, Recombination is required for efficient HIV-1 replication and the maintenance of viral genome integrity. *Nucleic Acids Res.* **46**, 10535–10545 (2018). doi: 10.1093/nar/gky910; pmid: 30307534
25. J. A. Briggs, T. Wilk, R. Welker, H. G. Kräusslich, S. D. Fuller, Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707–1715 (2003). doi: 10.1093/emboj/cdg143; pmid: 12660176
26. R. C. Burdick *et al*., HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5486–5493 (2020). doi: 10.1073/pnas.1920631117; pmid: 32094182
27. L. C. James, The HIV-1 capsid: More than just a delivery package. *Adv. Exp. Med. Biol.* **1215**, 69–83 (2019). doi: 10.1007/978-3-030-14741-9_4; pmid: 31317496
28. E. M. Campbell, T. J. Hope, HIV-1 capsid: The multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* **13**, 471–483 (2015). doi: 10.1038/nrmicro3503; pmid: 26179359
29. B. M. Forshey, U. von Schwedler, W. I. Sundquist, C. Aiken, Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002). doi: 10.1128/JVI.76.11.5667-5677.2002; pmid: 11991995
30. S. R. Yant *et al*., A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. *Nat. Med.* **25**, 1377–1384 (2019). doi: 10.1038/s41591-019-0560-x; pmid: 31501601
31. U. K. von Schwedler, K. M. Stray, J. E. Garrus, W. I. Sundquist, Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* **77**, 5439–5450 (2003). doi: 10.1128/JVI.77.9.5439-5450.2003; pmid: 12692245
32. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, Assembly properties of the human immunodeficiency virus type 1 CA protein. *J. Virol.* **78**, 2545–2552 (2004). doi: 10.1128/JVI.78.5.2545-2552.2004; pmid: 14963157
33. J. O. Link *et al*., Clinical targeting of HIV capsid protein with a long-acting small molecule. *Nature* **584**, 614–618 (2020). doi: 10.1038/s41586-020-2443-1; pmid: 32612233
34. C. W. Hooker, W. B. Lott, D. Harrich, Inhibitors of human immunodeficiency virus type 1 reverse transcriptase target distinct phases of early reverse transcription. *J. Virol.* **75**, 3095–3104 (2001). doi: 10.1128/JVI.75.7.3095-3104.2001; pmid: 11238836
35. O. Pornillos *et al*., X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282–1292 (2009). doi: 10.1016/j.cell.2009.04.063; pmid: 19523676
36. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011). doi: 10.1038/nature09640; pmid: 21248851
37. G. Zhao *et al*., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013). doi: 10.1038/nature12162; pmid: 23719463
38. C. L. Márquez *et al*., Kinetics of HIV-1 capsid uncoating revealed by single-molecule analysis. *eLife* **7**, e34772 (2018). doi: 10.7554/eLife.34772; pmid: 29877795
39. M. Novikova, Y. Zhang, E. O. Freed, K. Peng, Multiple roles of HIV-1 capsid during the virus replication cycle. *Virol. Sin.* **34**, 119–134 (2019). doi: 10.1007/s12250-019-00095-3; pmid: 31028522
40. R. A. Dick, D. L. Mallery, V. M. Vogt, L. C. James, IP6 regulation of HIV capsid assembly, stability, and uncoating. *Viruses* **10**, 640 (2018). doi: 10.3390/v10110640; pmid: 30445742
41. Z. Ambrose, C. Aiken, HIV-1 uncoating: Connection to nuclear entry and regulation by host proteins. *Virology* **454-455**, 371–379 (2014). doi: 10.1016/j.virol.2014.02.004; pmid: 24559861
42. L. Hilditch, G. J. Towers, A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Curr. Opin. Virol.* **4**, 32–36 (2014). doi: 10.1016/j.coviro.2013.11.003; pmid: 24525292
43. J. I. Mamede, G. C. Cianci, M. R. Anderson, T. J. Hope, Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7169–E7178 (2017). doi: 10.1073/pnas.1706245114; pmid: 28784755
44. S. Rankovic, J. Varadarajan, R. Ramalho, C. Aiken, I. Rousso, Reverse transcription mechanically initiates HIV-1 capsid disassembly. *J. Virol.* **91**, e00289-17 (2017). doi: 10.1128/JVI.00289-17; pmid: 28381579
45. A. Engelman, I. Oztop, N. Vandegraaff, N. K. Raghavendra, Quantitative analysis of HIV-1 preintegration complexes. *Methods* **47**, 283–290 (2009). doi: 10.1016/j.ymeth.2009.02.005; pmid: 19233280
46. F. Canducci *et al*., In vitro phenotypes to elvitegravir and dolutegravir in primary macrophages and lymphocytes of clonal recombinant viral variants selected in patients failing raltegravir. *J. Antimicrob. Chemother.* **68**, 2525–2532 (2013). doi: 10.1093/jac/dkt220; pmid: 23798668
47. Z. Temesgen, D. S. Siraj, Raltegravir: First in class HIV integrase inhibitor. *Ther. Clin. Risk Manag.* **4**, 493–500 (2008). doi: 10.2147/TCRM.S2268; pmid: 18728839
48. A. G. Holman, J. M. Coffin, Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6103–6107 (2005). doi: 10.1073/pnas.0501646102; pmid: 15802467
49. G. P. Wang, A. Ciuffi, J. Leipzig, C. C. Berry, F. D. Bushman, HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**, 1186–1194 (2007). doi: 10.1101/gr.6286907; pmid: 17545577
50. E. Serrao, A. Ballandras-Colas, P. Cherepanov, G. N. Maertens, A. N. Engelman, Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, 39 (2015). doi: 10.1186/s12977-015-0167-3; pmid: 25924943
51. C. M. Farnet, W. A. Haseltine, Circularization of human immunodeficiency virus type 1 DNA in vitro. *J. Virol.* **65**, 6942–6952 (1991). doi: 10.1128/JVI.65.12.6942-6952.1991; pmid: 1834863
52. H. Chen, A. Engelman, The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15270–15274 (1998). doi: 10.1073/pnas.95.26.15270; pmid: 9860958
53. N. Yan, P. Cherepanov, J. E. Daigle, A. Engelman, J. Lieberman, The SET complex acts as a barrier to autointegration of HIV-1. *PLOS Pathog.* **5**, e1000327 (2009). doi: 10.1371/journal.ppat.1000327; pmid: 19266025
54. B. F. Keele *et al*., Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7552–7557 (2008). doi: 10.1073/pnas.0802203105; pmid: 18490657
55. A. Rein, RNA Packaging in HIV. *Trends Microbiol.* **27**, 715–723 (2019). doi: 10.1016/j.tim.2019.04.003; pmid: 31085095
56. P. Bieniasz, A. Telesnitsky, Multiple, Switchable protein:RNA interactions regulate human immunodeficiency virus type 1 assembly. *Annu. Rev. Virol.* **5**, 165–183 (2018). doi: 10.1146/annurev-virology-092917-043448; pmid: 30048218
57. B. René, O. Mauffret, P. Fossé, Retroviral nucleocapsid proteins and DNA strand transfers. *Biochim Open* **7**, 10–25 (2018). doi: 10.1016/j.biopen.2018.07.001; pmid: 30109196
58. D. Castaño-Díez, M. Kudryashev, M. Arheit, H. Stahlberg, Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012). doi: 10.1016/j.jsb.2011.12.017; pmid: 22245546
59. K. A. Skorupka *et al*., Hierarchical assembly governs TRIM5α recognition of HIV-1 and retroviral capsids. *Sci. Adv.* **5**, eaaw3631 (2019). doi: 10.1126/sciadv.aaw3631; pmid: 31807695
60. K. Qu *et al*., Structure and architecture of immature and mature murine leukemia virus capsids. *Proc. Natl. Acad.*

Sci. U.S.A. **115**, E11751–E11760 (2018). doi: 10.1073/pnas.1811580115; pmid: 30478053
61. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
62. A. Morin *et al*., Collaboration gets the most out of software. *eLife* **2**, e01456 (2013). doi: 10.7554/eLife.01456; pmid: 24040512
63. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
64. J. T. Robinson *et al*., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011). doi: 10.1038/nbt.1754; pmid: 21221095
65. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014). doi: 10.1093/bioinformatics/btu393; pmid: 24930139
66. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
67. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004). doi: 10.1101/gr.849004; pmid: 15173120
68. T. D. Schneider, R. M. Stephens, Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990). doi: 10.1093/nar/18.20.6097; pmid: 2172928
69. D. E. Christensen, B. K. Ganser-Pornillos, J. S. Johnson, O. Pornillos, W. I. Sundquist, immunoviraldesign, immunoviraldesign/HIV_in_vitro_integration: Published_release. Zenodo (2020). doi: 10.5281/zenodo.3966034

ACKNOWLEDGMENTS

We thank B. Lohman in the Genomics and Bioinformatics Analysis group at the Huntsman Cancer Institute for performing integration site sequencing alignments, Y. Wan for assistance with large-scale virus production, M. Pereira for measuring the infectivity of pseudotyped ΔEnvSG3 virions, K. Dryden for expert EM assistance, Gilead Sciences for providing GS-CA1 inhibitor, and J. Iwasa for creating the overview figure for the print summary. Oligonucleotides were synthesized by the University of Utah DNA/Peptide Core Facility, sequencing was performed at the University of Utah DNA sequencing Core Facility, and EM data were collected at the University of Virginia Molecular Electron Microscopy Core facility, which is supported in part by the School of Medicine and built with NIH grant G20-RR31199. The Titan Krios (SIG S10-RR025067), Falcon II/3EC direct detector (SIG S10-OD018149), and K3/GIF (U24-GM116790) were purchased in part or in full with the indicated NIH grants. **Funding**: We acknowledge funding from NIH P50 AI150464 (J.S.J., O.P., and W.I.S.) and the University of Virginia School of Medicine Pinn Scholar Program (O.P.). **Author contributions**: Conceptualization: D.E.C., B.K.G.-P., J.S.J., O.P., and W.I.S. Investigation: D.E.C.,. B.K.G.-P., J.S.J., and O.P. Supervision: O.P. and W.I.S. Writing, original draft: W.I.S. Writing, review and editing: All authors. **Competing interests**: The authors declare no competing interests. **Data and materials availability**: All plasmids used in this study are available at Addgene (table S2). 3′ integration site deep sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) database under accession no. PRJNA649355. Python scripts used for sequence analyses are available from an archived GitHub repository (*69*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/eabc8420/suppl/DC1
Figs. S1 to S9
Tables S1 and S2
Movies S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

16 May 2020; accepted 31 July 2020
10.1126/science.abc8420

RESEARCH ARTICLE SUMMARY

VIRAL INFECTIONS

Neutrophilic inflammation in the respiratory mucosa predisposes to RSV infection

Maximillian S. Habibi*, Ryan S. Thwaites*, Meiping Chang, Agnieszka Jozwik, Allan Paras, Freja Kirsebom, Augusto Varese, Amber Owen, Leah Cuthbertson, Phillip James, Tanushree Tunstall, David Nickle, Trevor T. Hansel, Miriam F. Moffatt, Cecilia Johansson‡, Christopher Chiu†‡, Peter J. M. Openshaw†‡

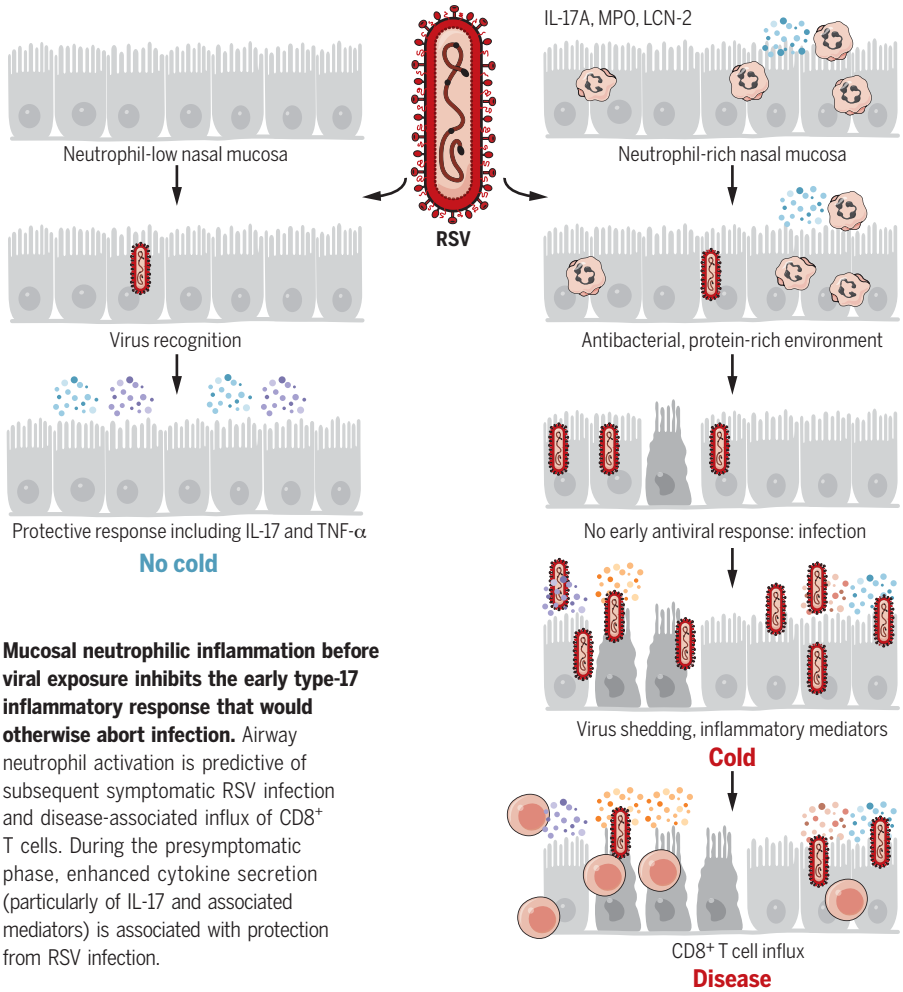
INTRODUCTION: Even with intimate exposure to a virus, some people fail to become infected. Variable transmission partly depends on the dose and duration of exposure but is also governed by the immune status of the host, such as the presence of specific protective antibodies or T cells. However, for some infections, the reasons for erratic transmission are largely unknown. For example, respiratory syncytial virus (RSV) can repeatedly reinfect individuals throughout their lives despite the presence of specific immunity. Additionally, antibodies and T cells have limited efficacy against newly emergent pathogens with pandemic potential. However, the intrinsic and innate mechanisms underlying protection when people are exposed to these viruses are poorly understood.

RATIONALE: We reasoned that the prior state of the respiratory mucosa's innate defenses may contribute to the variable outcome of RSV inoculation. By performing experimental challenge of adult volunteers, we were able to measure variations in the status of the nasal mucosa before inoculation and in mucosal responses during the presymptomatic phase of infection. Neither of these phases is easily observable during natural spontaneous transmission. Our observations could then be validated using specific interventional studies in a well-established mouse model of RSV infection.

RESULTS: After nasal administration of RSV, 57% of inoculated volunteers became infected. The uptake of infection was poorly explained by specific B or T cell immunity. However, transcriptomic profiling of the nasal tissue before inoculation demonstrated a neutrophilic inflammatory signal in those destined to develop symptomatic infection, and this was associated with suppression of an early interleukin-17 (IL-17)–dominated immune response during the presymptomatic period. This was followed by symptomatic infection associated with the expression of proinflammatory cytokines. By contrast, those who resisted infection showed a transient boost in mucosal markers of innate immune activation for several days after viral administration and no subsequent viral replication. In mice, chemokine-induced neutrophil recruitment to the airway before viral exposure transiently enhanced viral replication immediately after inoculation. As with human subjects, this was associated with decreased expression of some innate mediators in the respiratory tract, which was then followed by enhanced disease driven by $CD8^+$ T cells.

CONCLUSION: Although reinfection with a respiratory virus can be partially explained by the waning of antibody titers and T cell numbers, neutrophilic inflammation in the airway at the time of pathogen exposure also predisposes individuals to symptomatic infection. After exposure, the response of the mucosa diverges: A mild and transient increase in nasal inflammatory mediators is accompanied by termination of viral infection, whereas failure to mount this response is followed by viral success.

The state of innate immune preparedness of the respiratory mucosa is thus a major determinant of susceptibility to viral challenge. Our results show the importance of understanding the mucosal microenvironment in studies of respiratory infection and highlight targets for local intervention, which may enhance protection against a range of respiratory pathogens. ■



Mucosal neutrophilic inflammation before viral exposure inhibits the early type-17 inflammatory response that would otherwise abort infection. Airway neutrophil activation is predictive of subsequent symptomatic RSV infection and disease-associated influx of $CD8^+$ T cells. During the presymptomatic phase, enhanced cytokine secretion (particularly of IL-17 and associated mediators) is associated with protection from RSV infection.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: p.openshaw@imperial.ac.uk (P.J.M.O.); c.chiu@imperial.ac.uk (C.C.); c.johansson@imperial.ac.uk (C.J.)
Cite this article as M. S. Habibi *et al.*, *Science* **370**, eaba9301 (2020). DOI: 10.1126/science.aba9301

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba9301

GRAPHICS: N. CARY/*SCIENCE*

RESEARCH ARTICLE

VIRAL INFECTIONS

Neutrophilic inflammation in the respiratory mucosa predisposes to RSV infection

Maximillian S. Habibi[1]*, Ryan S. Thwaites[1]*, Meiping Chang[2], Agnieszka Jozwik[1], Allan Paras[1], Freja Kirsebom[1], Augusto Varese[1], Amber Owen[1], Leah Cuthbertson[1], Phillip James[1], Tanushree Tunstall[1], David Nickle[3], Trevor T. Hansel[1], Miriam F. Moffatt[1], Cecilia Johansson[1]‡, Christopher Chiu[4]†‡, Peter J. M. Openshaw[1]†‡

The variable outcome of viral exposure is only partially explained by known factors. We administered respiratory syncytial virus (RSV) to 58 volunteers, of whom 57% became infected. Mucosal neutrophil activation before exposure was highly predictive of symptomatic RSV disease. This was associated with a rapid, presymptomatic decline in mucosal interleukin-17A (IL-17A) and other mediators. Conversely, those who resisted infection showed presymptomatic activation of IL-17– and tumor necrosis factor–related pathways. Vulnerability to infection was not associated with baseline microbiome but was reproduced in mice by preinfection chemokine-driven airway recruitment of neutrophils, which caused enhanced disease mediated by pulmonary $CD8^+$ T cell infiltration. Thus, mucosal neutrophilic inflammation at the time of RSV exposure enhances susceptibility, revealing dynamic, time-dependent local immune responses before symptom onset and explaining the as-yet unpredictable outcomes of pathogen exposure.

The respiratory tract is by necessity an open portal, and thus is highly vulnerable to invasion by many pathogens. A complex system of defenses counters this vulnerability, but current knowledge of what governs susceptibility fails to fully explain the erratic transmission of infectious agents (*1*, *2*). For some infections, antibodies in the respiratory mucosa are clearly protective (*3*). Infectious agents can also be blocked by antimicrobial peptides (*4*), entrapment in mucus (*5*), and the downstream effects of pattern recognition receptor activation (*1*, *2*, *6*), as well as by the direct inactivation by immune cells (*6*). The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has brought into sharp focus the need for a greater understanding of the role of mucosal innate immunity in protection against respiratory infection. Better understanding of such defenses may offer ways to prevent or modulate viral disease, even that caused by novel and emerging pathogens. Among innate immune cells, neutrophils are classically ascribed roles in defense against bacteria and fungi, but their role in antiviral responses is less clear. Although neutrophils are sometimes found in the lungs of patients with severe viral infections and have a demonstrable antiviral role in animal models, information concerning how they contribute to defense and disease in human respiratory viral infections is largely lacking (*7*).

Respiratory syncytial virus (RSV) is the leading cause of infant hospitalization worldwide, infecting ~34 million children each year (*8*) and contributing prominently to morbidity and mortality in elderly and immunosuppressed adults (*9*, *10*). Although RSV infection is essentially confined to the respiratory epithelium, the resultant neutrophilic lung inflammation can be life-threatening, particularly in at-risk children (*11*, *12*). Although prophylaxis with palivizumab (a monoclonal anti-RSV antibody) can prevent hospitalization of premature infants (*13*), no specific therapeutic intervention is currently available and no vaccine has yet been licensed that protects against RSV (*14*). This is despite the existence of several promising candidates (*15*), including enhancement of placental anti-RSV immunoglobulin G (IgG) transfer by maternal immunization. Accordingly, there is a pressing need to identify additional interventions that prevent RSV and other respiratory viral diseases.

Animal models of respiratory viral infection provide useful mechanistic insights but do not fully recapitulate human disease (*6*). Observational studies of hospitalized children and high-risk adults are limited by the heterogeneity of populations and disease severity and by an inability to dynamically assess the status of host immunity—specifically, the mucosa before and immediately after viral exposure. Experimental infection of human volunteers therefore provides an unparalleled opportunity in a controlled setting to identify preexisting immune factors and presymptomatic responses that correlate with protection and disease (*16*, *17*).

All adults have RSV-specific antibodies, reflecting multiple rounds of infection throughout the course of their lives. Nevertheless, symptomatic reinfection occurs repeatedly even in healthy people and is associated with unusually short-lived humoral and cellular memory responses (*3*). Such defective defense against reinfection is also observed after some other respiratory viral infections, including coronaviruses (*18*). Furthermore, variations in antibody levels do not accurately predict susceptibility to RSV infection in experimental studies of volunteers: Those with the lowest antibody levels are only modestly more susceptible to viral challenge (*3*, *19*). Although mucosal IgA, circulating IgG (*3*), and resident memory $CD8^+$ T cells (*20*) show some association with resistance to infection and reduced disease severity, these factors do not completely explain who resists and who succumbs. Additionally, although RSV viral burden approximately correlates with symptoms in experimentally infected adults (*16*) and moderately ill infants (*21*, *22*), viral load may be paradoxically reduced in the most severely affected children (*23*).

How the very early events after viral exposure influence clinical outcomes has not been studied in either adults or children. By studying the respiratory mucosa before, during, and after human RSV challenge, we found that prior neutrophil activation in the upper respiratory tract predisposed individuals to symptomatic viral infection. The neutrophilic mucosal environment was strongly associated with a reduction in antiviral mucosal inflammatory responses immediately after viral exposure that was followed by the onset of disease. By contrast, a protective response characterized by rapid activation of type-17 inflammation was observed in those resisting infection. Moreover, studies in mice showed that recruitment of airway neutrophils before viral administration enhanced the cytotoxic $CD8^+$ T cell response and disease severity.

Results

Characterization of the time course of responses to human RSV challenge

Fifty-eight healthy adult volunteers were inoculated with RSV Memphis 37. Nasal sampling was performed at baseline (7 to 14 days before inoculation) and then repeatedly up to 14 days postinoculation (dpi) (Fig. 1A). Symptoms were quantified using a previously validated self-reported symptom scale (*24*). Twenty-three participants developed polymerase chain reaction (PCR)–positive RSV infection with symptoms of upper respiratory tract disease (the "Cold" group), whereas 25 showed no evidence of RSV infection and did not develop

[1]National Heart and Lung Institute, Imperial College London, London, UK. [2]Merck & Co., Inc., Kenilworth, NJ, USA. [3]Genetics & Pharmacogenomics, Department of Translational Medicine, Merck & Co., Inc., Boston, MA, USA. [4]Department of Infectious Disease, Imperial College London, London, UK.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: p.openshaw@imperial.ac.uk (P.J.M.O.); c.chiu@imperial.ac.uk (C.C.); c.johansson@imperial.ac.uk (C.J.)

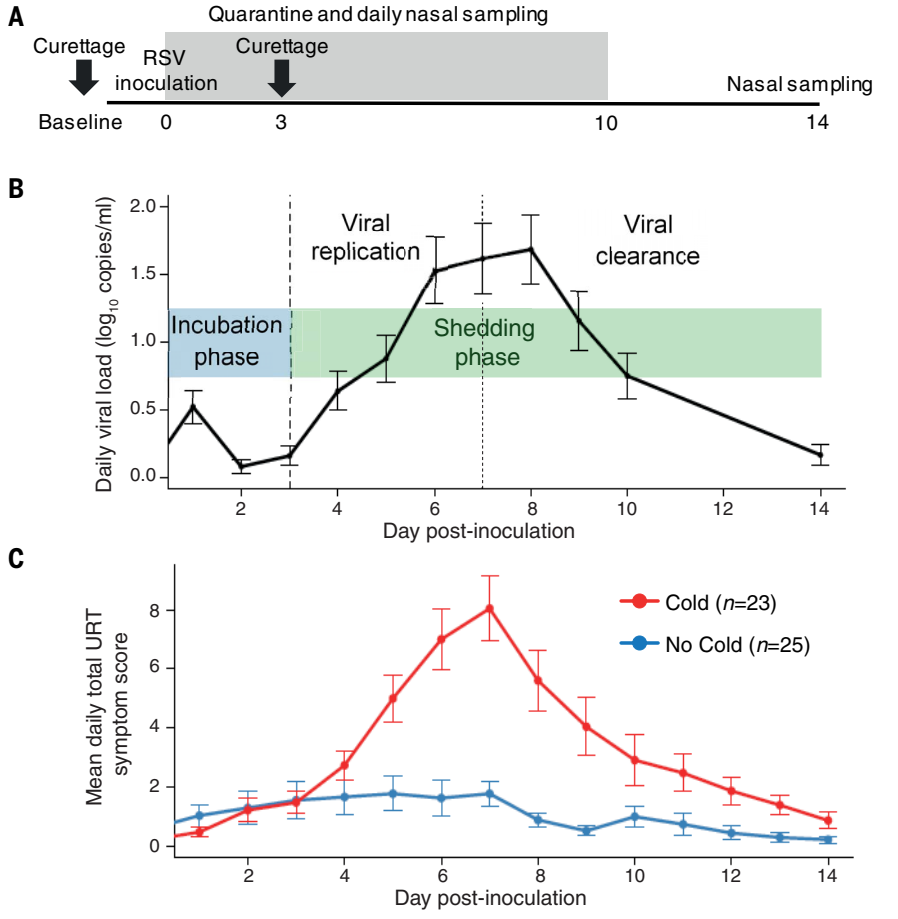


Fig. 1. Kinetics of viral replication and symptoms correlate closely, with no viral replication or symptoms evident until 3 dpi. (**A**) Participants were inoculated with RSV at day 0. Nasal wash and nasosorption samples were taken daily during quarantine (up to 10 dpi) and at 14 dpi. Nasal curettage was performed at baseline (7 to 14 days before inoculation) and at 3 dpi. (**B**) Daily viral load, measured by qPCR of nasal wash, in volunteers who developed symptomatic RSV infections (Cold, *n* = 23). (**C**) Daily upper respiratory tract (URT) symptom scores in the Cold (red, *n* = 23) and No Cold (blue, *n* = 25) groups. Data in (B) and (C) are shown as the means ± SEM.

symptoms (the "No Cold" group). Ten participants had PCR evidence of RSV infection but had self-reported symptom scores that did not reach the threshold required for the Cold group. This indeterminate group was excluded from subsequent analyses, which focused on identifying factors associated with the most distinct outcomes. No significant demographic differences were observed between the Cold and No Cold groups (table S1).

In infected volunteers, viral load [determined by quantitative PCR (qPCR) from daily nasal washes] broadly indicated three phases: (1) a presymptomatic incubation phase during the first 3 dpi when virus was undetectable; (2) a viral replication phase from 3 to 7 dpi; and (3) a viral clearance phase from 8 dpi [previously shown to be associated with the rise in virus-specific $CD8^+$ T cells in both the blood and lower airway (*20*)] (Fig. 1B). Low-level detection of virus at 1 dpi probably represented residual viral inoculum. Severity of the resultant upper respiratory tract disease (measured by a self-reported symptom score) peaked at 7 dpi, around the same time as viral load (Fig. 1C). In agreement with previous reports (*16*), total symptom scores positively correlated with total viral shedding throughout the study duration ($P < 0.001$, fig. S1) despite the known shortcomings of viral load measurement and the variability of symptom quantification.

Susceptibility to RSV infection is associated with airway neutrophil activation

Human challenge studies are exceptional in allowing the investigation of events before symptom onset, enabling us to examine the influence of immune activation in the airway at baseline and during the early presymptomatic period on the course of infection. In agreement with our previous report (*3*), anti-RSV IgA titers in the nose at baseline were higher in the No Cold group (fig. S2, A and B). However, titers showed a substantial overlap with the Cold group, suggesting that nasal IgA alone did not adequately explain susceptibility unless it was at extremely low or high levels. Additionally, the frequency of RSV-specific $CD8^+$ T cells in the lung at baseline did not distinguish between infection (Cold) or protection (No Cold) (fig. S2C), as we have previously reported (*20*). Thus, correlates of protection against RSV are likely multifactorial, with no single dominant protective mechanism yet defined.

To enable the discovery of previously unanticipated factors associated with susceptibility to infection, RNA sequencing (RNA-seq) was performed on nasal tissue collected from Cold and No Cold group volunteers before inoculation. This approach identified 80 differentially expressed genes (DEGs) that distinguished the two divergent outcomes [adjusted *P* value (P_{adj}) < 0.05; Fig. 2A and table S2]. Seventy-three (91%) DEGs were significantly higher in the Cold group, whereas a smaller cluster (7/80, 9%) was elevated in the No Cold group. For pathway enrichment analyses, a looser cutoff ($P_{adj} < 0.1$) was applied, resulting in inclusion of a further 155 DEGs (table S2). Biological processes associated with neutrophil activation made up the most significantly enriched pathways at baseline in those susceptible to infection (Fig. 2B and table S3), with many DEGs also associated with hydrolase and peptidase activity (Fig. 2C and table S4). Furthermore, weighted gene coexpression network analysis (WGCNA) identified a cluster of highly correlated genes enriched for genes associated with neutrophil activity, including *CDA*, *IGF2R*, *ITGAX*, *UNC13D*, *SIGLEC14*, *SIRPA*, *PREX1*, and *MMP9* (Fig. 2D). In agreement with these transcriptomic data, nasal protein levels of the neutrophil-associated mediators myeloperoxidase (MPO), lipocalin-2/neutrophil gelatinase-associated lipocalin (LCN-2), and interleukin-17A (IL-17A) were significantly higher in the Cold group relative to the No Cold group at day 0 ($P = 0.010$, $P = 0.017$, and $P = 0.039$, respectively; Fig. 2E and table S5). There was no difference in other cytokines or chemokines (table S5), but with IL-17 known to induce neutrophil activation and chemotaxis, these data support the hypothesis that greater airway neutrophil activity at the time of virus exposure increases susceptibility to symptomatic RSV infection (*25*, *26*). No correlation was observed between baseline nasal antibody titers and neutrophil mediator levels, including MPO (fig. S3), suggesting that these were independent correlates of susceptibility.

Early immune activation prevents symptomatic RSV infection

To investigate the relationship between airway neutrophil activity and subsequent mucosal inflammatory responses, nasal tissue samples

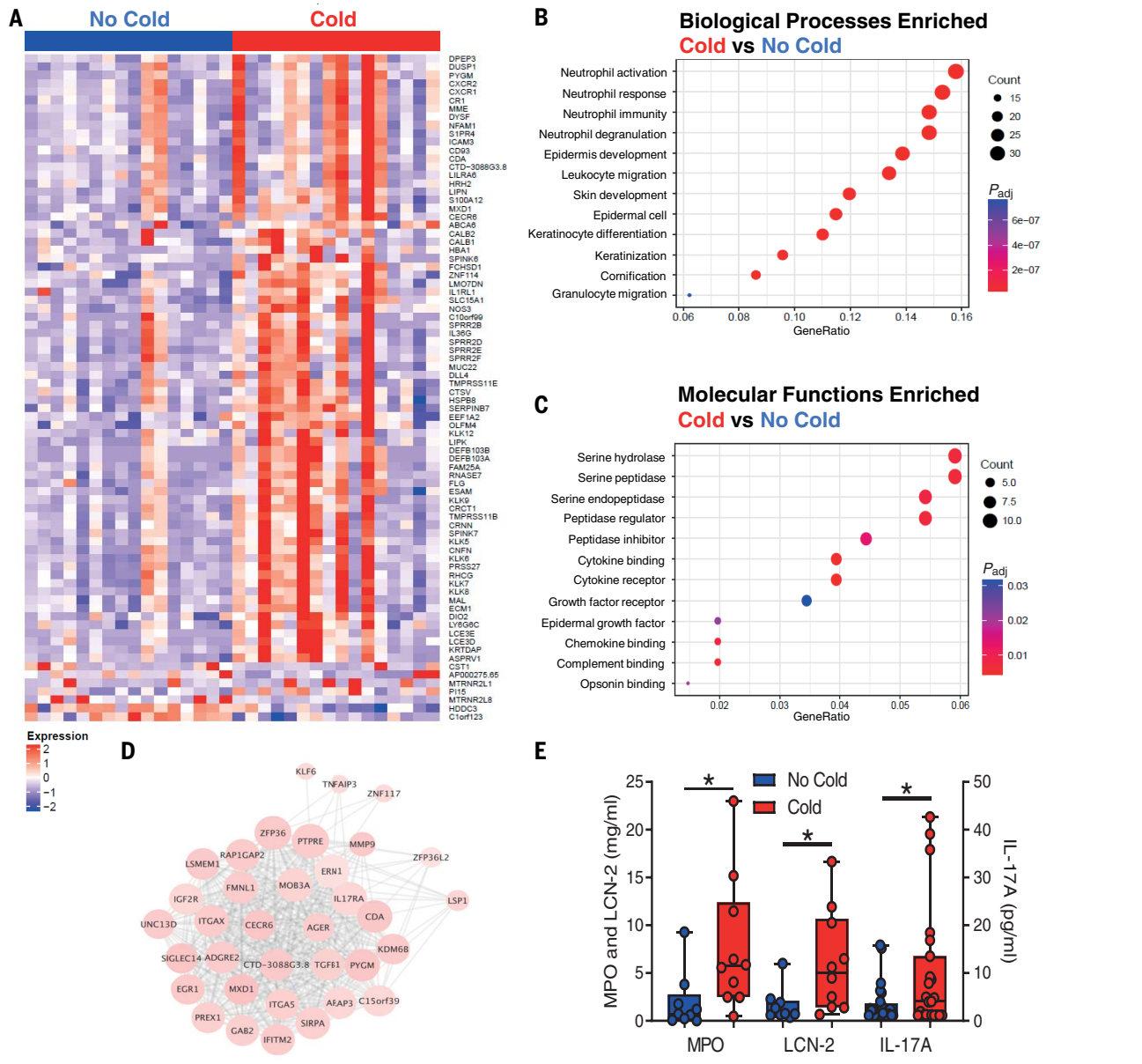


Fig. 2. Preexposure airway neutrophil activation is associated with susceptibility to RSV infection. (**A**) DEGs (P_{adj} < 0.05; n = 80) were determined in baseline (7 to 14 days before inoculation) nasal curettage samples between Cold and No Cold groups by RNA-seq. (**B**) GO molecular functions enrichment analysis of DEGs (P_{adj} < 0.1, n = 235). (**C**) GO biological processes enrichment analysis of DEGs (P_{adj} < 0.1, n = 235). (**D**) WGCNA cluster of DEGs (P_{adj} < 0.1, n = 235) enriched for those with annotations of neutrophil association. (**E**) Day 0 levels of MPO, LCN-2, and IL-17A in nasosorption samples between the No Cold (n = 10-25) and Cold (n = 9-23) groups. DEGs in (A) to (D) were determined by Wald tests using DESeq2. Data in (E) are shown as median and interquartiles with minimum and maximum values and were analyzed by Mann–Whitney U tests. *P < 0.05.

collected at 3 dpi (before the phase of concerted net viral replication) were analyzed by RNA-seq and compared with matched pre-inoculation samples. Sixteen volunteers from each of the Cold and No Cold groups yielded RNA suitable for sequencing and were included in the analysis. This analysis identified 87 DEGs (P < 0.01, $\log^2$ fold change > 0.5) in the No Cold group (Fig. 3A and table S6) and 77 DEGs in the Cold group (Fig. 3B and table S7). In both the Cold and No Cold groups, most DEGs were up-regulated after inoculation (as shown by the mean expression values for each gene and K-means clustering). However, only two genes (*ITLN1* and *JUN*) overlapped between both groups, emphasizing the divergence of local responses associated with these two opposing outcomes of viral exposure (Fig. 3C).

Next, the functional pathways to which DEGs associated with the No Cold group at day 3 significantly contributed were tested by enrichment analysis using the gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathways. GO biological processes analysis for the Cold group showed significant enrichment of several processes but none specifically related to immune responses (table S8). By contrast, in the No Cold group, GO biological processes analysis yielded several immune-related pathways that were significantly enriched. These could be categorized into three areas: (1) epidermis/skin development and bacterial defense; (2) humoral immunity, including the *JCHAIN* required for secretory IgA formation; and (3) leukocyte/

granulocyte migration and chemotaxis (table S9). Additionally, the DEGs in the No Cold group showed the enrichment of molecular functions for protease, cytokine, and chemokine activities (Fig. 3D and table S10). KEGG analysis also revealed several pathways related to cytokine and chemokine signaling. Of these, the "IL-17 signaling pathway" was the most significantly enriched (Fig. 3E and table S11), comprising chemokines *CXCL1/2/3/6*, the transcription factor *JUN*, the major gel-forming mucin *MUC5B*, and the antimicrobial protein *S100A7*, which is induced by IL-17 and important in protection against Gram-negative bacteria (*27*). Other closely related KEGG terms were also enriched, including "Cytokine–cytokine receptor interaction" and "TNF signaling pathway." No KEGG pathways were significantly enriched in the Cold group.

These findings show that the activation of an early inflammatory response is associated with protection from symptomatic RSV infection. This early protective inflammatory response is characterized by signatures of leukocyte migration and cytokine and chemokine signaling. By contrast, such inflammatory responses were not observed in volunteers who subsequently developed RSV disease.

Early RSV clearance is associated with type-17 immunity

To define the core transcriptional signatures of coexpressed genes without biasing by predetermined or putative biological function knowledge, we next applied WGCNA to

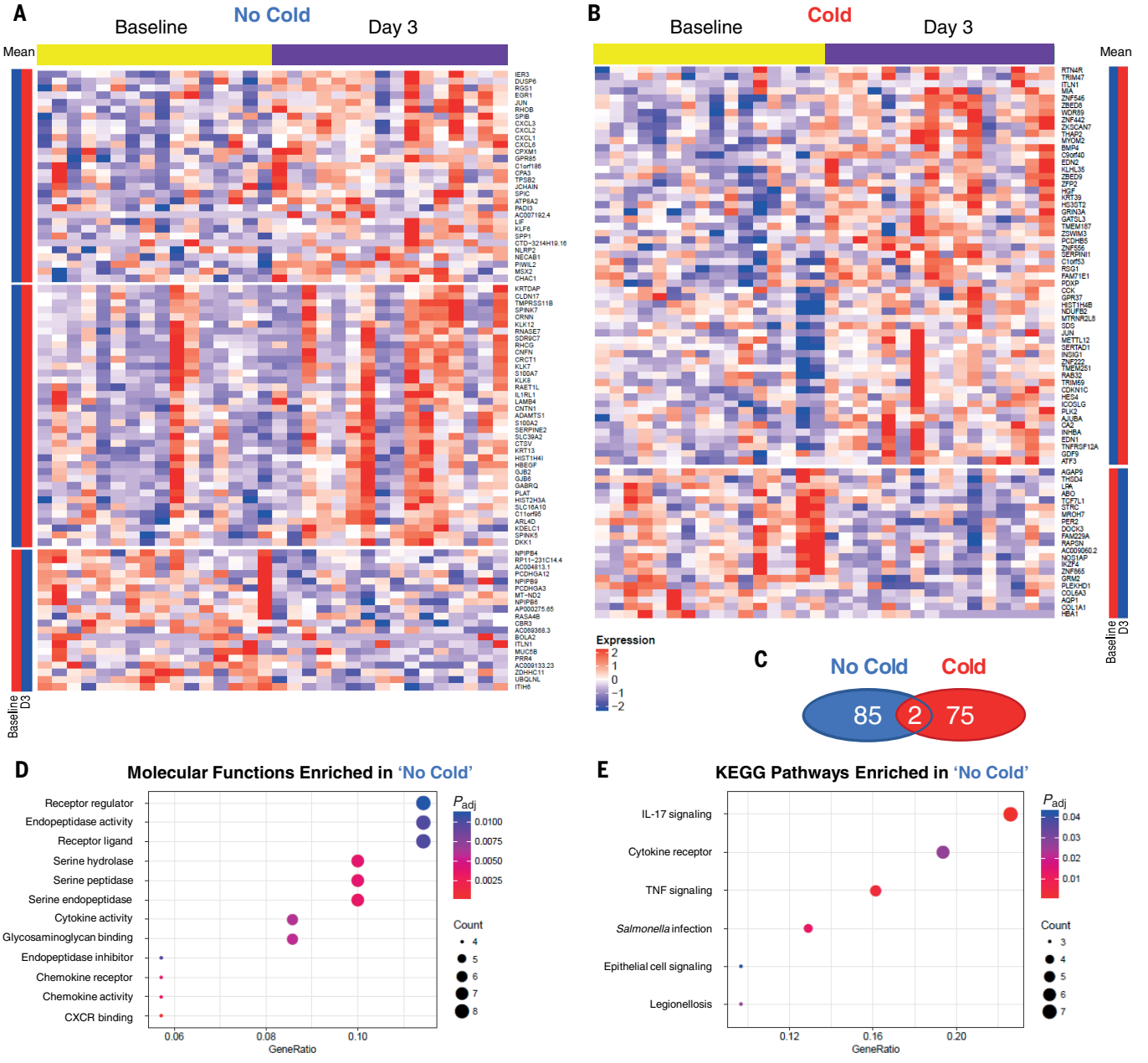


Fig. 3. Early presymptomatic viral clearance is associated with activation of IL-17 signaling. DEGs were determined by RNA-seq of nasal curettage samples at 3 dpi relative to baseline (7 to 14 days before inoculation) in the No Cold (n = 16) and Cold (n = 16) groups. Heatmaps show (**A**) DEGs in the No Cold group (n = 87) and (**B**) DEGs in the Cold group (n = 77). (**C**) Venn diagram of the number of overlapping DEGs in the Cold and No Cold groups. DEGs from the No Cold group were analyzed for enrichment of (**D**) GO molecular functions and (**E**) KEGG pathways. DEGs were defined as transcripts with P_{adj} < 0.01 and $\log^2$ fold change >0.5.

establish interrelated DEG networks associated with protection in No Cold individuals. Three gene clusters (depicted and named "brown," "turquoise," and "blue") were identified at 3 dpi relative to baseline (Fig. 4A). The brown module mainly contained genes encoding peptidase and hydrolase activity (Fig. 4, B and C, and table S12). This was also true of the turquoise module, which clustered closely with the brown module (Fig. 4, D and E, and table S13). However, the blue module was highly enriched in IL-17 pathway components and closely corresponded to the DEGs featured in the KEGG "IL-17 signaling pathway" (Fig. 3E), namely *CXCL1/2/3/6* and *JUN* (Fig. 4F). Indeed, GO molecular function analysis of the blue module identified enrichment of cytokine and chemokine activity and signaling (Fig. 4G and table S14). Thus, the blue module represents a signature of type-17 inflammatory gene expression induced on RSV exposure, which is associated with the prevention of infection.

To validate these findings, cytokines and chemokines in the nasal mucosal lining fluid were quantified over time. In the Cold group, a limited response was detected until 3 dpi, after which the expression of antiviral mediators [interferon-α (IFN-α), IFN-γ, CXCL10, and IL-15] substantially increased, with maximal detection at 7 dpi (Fig. 5A), which is consistent with the peak of symptoms and viral load (Fig. 1, B and C). The chemokines CCL3 (MIP-1α) and CCL5 (RANTES), the anti-inflammatory cytokine IL-10, and tumor necrosis factor-α (TNF-α) all followed similar kinetics (Fig. 5B). The levels of all mediators in the Cold group declined during the phase of viral clearance (7 to 14 dpi). Cytokines and chemokines are well recognized to be closely associated with symptoms in a variety of inflammatory conditions, including the response to infection (*28*, *29*). Indeed, within the Cold group, cumulative viral load and inflammatory mediator levels over the study duration were closely correlated (fig. S4), emphasizing their role in proinflammatory responses and disease triggered by infection (*16*). However, in the No Cold group, despite the absence of symptoms before 4 dpi, several cytokines (including IL-1β, IL-6, IL-17A, and TNF-α) underwent a rapid burst of expression

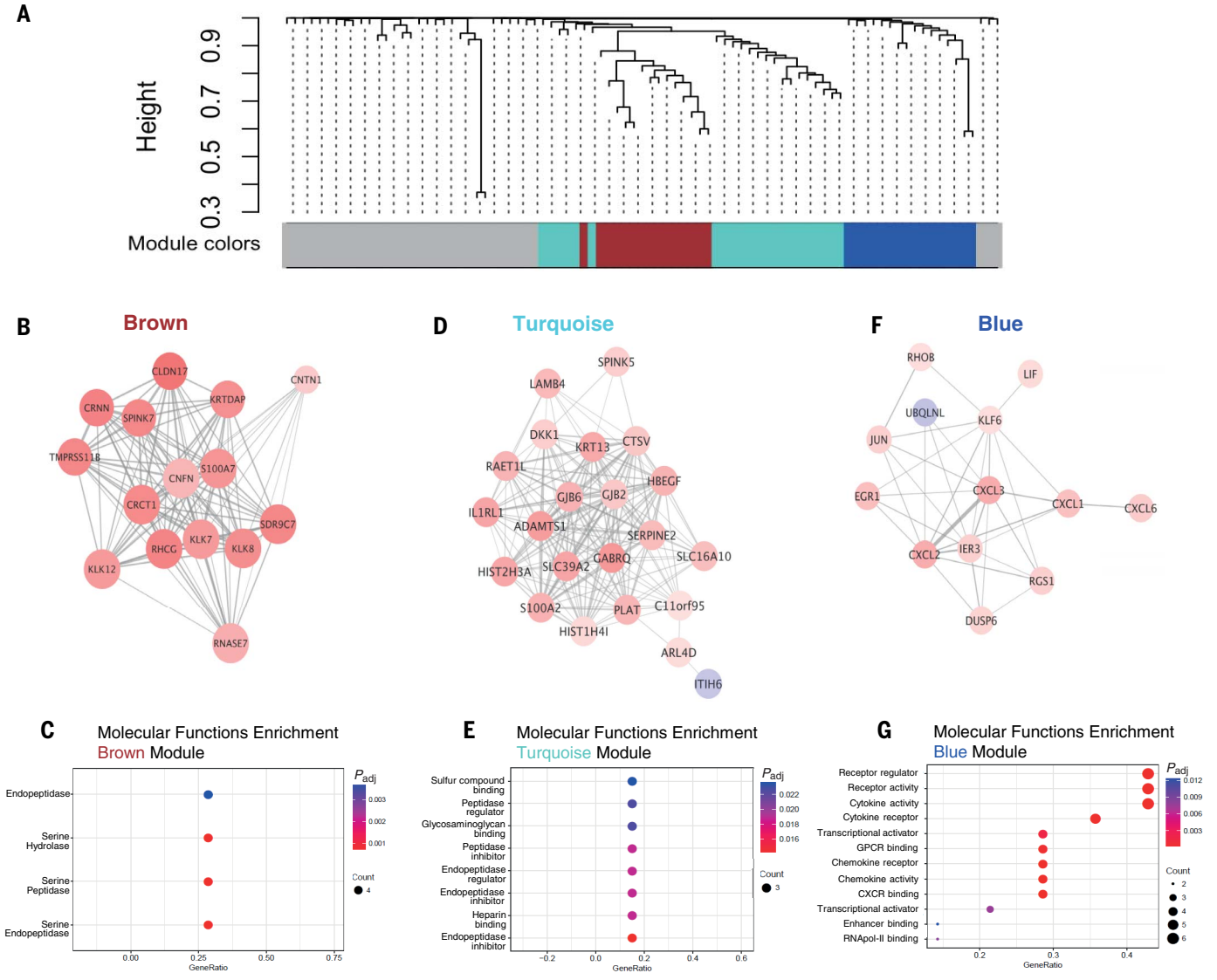
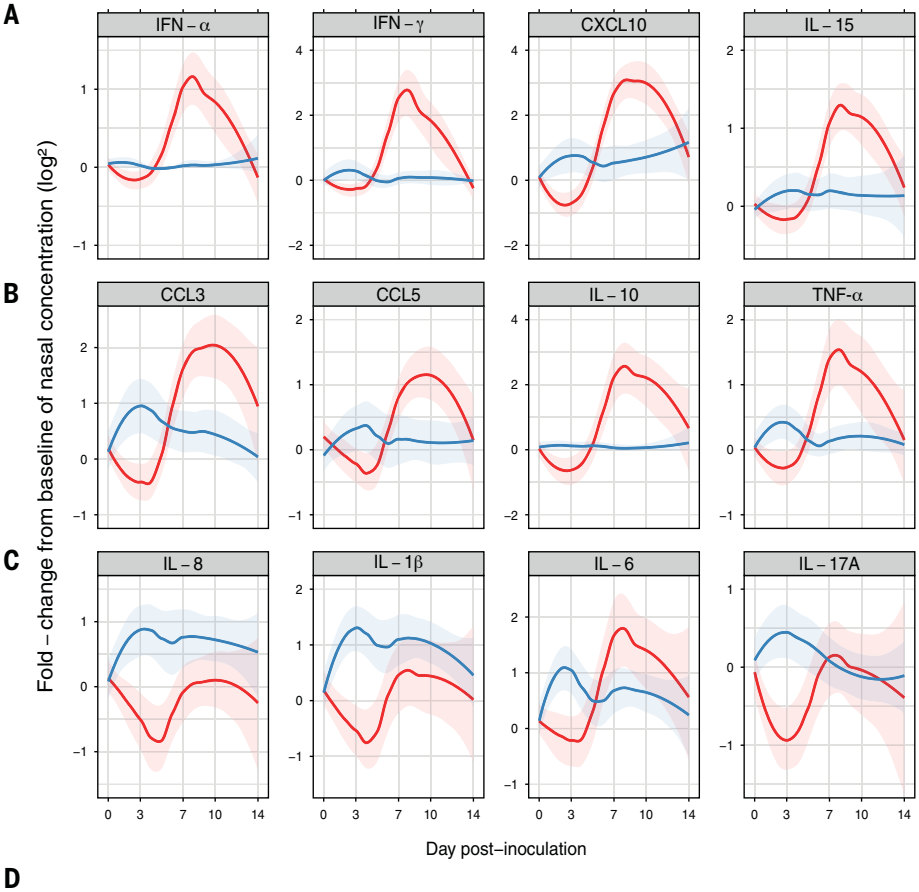


Fig. 4. Network analysis reveals a module of correlated DEGs enriched for IL-17 signaling in presymptomatic individuals. WGCNA of the differentially expressed genes (n = 87) was performed in the No Cold group at 3 dpi relative to baseline. (**A**) Classification of WGCNA modules. Brown module: (**B**) network map and (**C**) molecular function enrichment. Turquoise module: (**D**) network map and (**E**) molecular function enrichment. Blue module: (**F**) network map and (**G**) molecular function enrichment. DEGs were defined as transcripts with $P < 0.01$ and $\log^2$ fold change >0.5.



D

Mediator	IFN-α	IFN-γ	CXCL10	IL-15	CCL3	CCL5	IL-10	TNF-α	IL-8	IL-1β	IL-6	IL-17A
P-value	0.20	0.22	0.07	0.41	0.10	0.92	0.10	0.12	0.10	0.04*	0.05*	0.03*
Odds-ratio (95% CI)	0.49 (0.10, 1.06)	0.83 (0.57, 1.06)	0.88 (0.74, 1.00)	0.88 (0.63, 1.19)	0.87 (0.71, 1.01)	1.01 (0.83, 1.23)	0.78 (0.52, 0.99)	0.82 (0.61, 1.03)	0.85 (0.69, 1.01)	0.86 (0.73, 0.98)	0.82 (0.65, 0.97)	0.80 (0.63, 0.95)

Fig. 5. Early presymptomatic secretion of intranasal IL-1β, IL-6, and IL-17A is associated with protection from RSV infection. (**A** to **C**) Soluble protein cytokine and chemokine mediator levels from nasosorption samples were determined on each study day and expressed as LOESS plots between Cold (red) and No Cold (blue) groups of $\log^2$ fold-change levels normalized to day 0. The 95% confidence intervals of LOESS curves are denoted by shaded areas. (**D**) Logistic regression analysis of mediator responses during 1 to 3 dpi measured at the protein level from nasosorption samples.

during the incubation phase shortly after virus inoculation (Fig. 5, B and C). Logistic regression analysis showed statistically significant associations between the induction of IL-1β, IL-6, and IL-17A (P = 0.04, 0.05, and 0.03, respectively; Fig. 5D) in the first 3 days after RSV exposure and protection from developing symptoms. After 3 dpi, mediators returned to baseline levels in No Cold volunteers, who did not develop the later response seen in the Cold group. Conversely, in the Cold group, almost all mediators declined relative to day 0 levels during the incubation phase despite the presence of virus (Fig. 5, A to C). Thus, a local type-17 inflammatory response early after pathogen exposure prevents the development of a symptomatic infection. Failure to mount a protective inflammatory response rapidly after inoculation is itself associated with airway neutrophil activity at the time of exposure, suggesting that neutrophilic inflammation may disrupt this early protective response.

The airway microbiome is not associated with susceptibility to infection

Neutrophils are classically ascribed antibacterial and antifungal functions, and recent studies have shown perturbations of the airway bacterial microbiome in children hospitalized with RSV (*30*). To test for associations between the preinfection airway microbiome and susceptibility to RSV infection, 16*S* ribosomal RNA from nasopharyngeal bacteria was quantified and sequenced. Despite differential neutrophil activation in the nasal mucosa seen by multiple measures, no significant differences in bacterial load were observed between the Cold and No Cold groups at baseline (fig. S5A). Moreover, 16*S* sequencing demonstrated similar baseline diversity and community composition in these groups (richness, P = 0.948; Shannon–Weiner, P = 0.263; Simpson's, P = 0.166, Bray–Curtis dissimilarity, R^2 = 0.055, P = 0.087; fig. S5B). Thus, there are no significant associations between the baseline airway microbiome and nasal neutrophilic inflammation before viral exposure and thus susceptibility to RSV infection.

Elevated airway neutrophils in the murine lung at the time of exposure enhance RSV disease through pulmonary cytotoxic T cells

To investigate the role of neutrophilic inflammation on early and late outcomes in respiratory viral infection, we modeled the impact of increased pulmonary neutrophils at the time of RSV challenge in mice. C57BL/6 mice were treated intranasally with recombinant CXCL1 (rCXCL1) or phosphate-buffered saline (PBS). Administration of rCXCL1 caused short-lived recruitment of neutrophils to the bronchoalveolar lavage (BAL) and lung parenchyma within 12 hours (Fig. 6A; gating strategies are shown in fig. S6), resulting in elevated levels of the neutrophil mediators MMP-9, MPO, and neutrophil elastase in BAL (Fig. 6B); this was analogous to human volunteers in the Cold group. The effect of rCXCL1 treatment was restricted to neutrophils and had no impact on either alveolar macrophage numbers in the airway or monocyte abundance in the lung (fig. S7A). Furthermore, no differences were observed in cytokine (fig. S7B) or chemokine (fig. S7C) levels in the lung before virus challenge after rCXCL1.

As in human challenge, increased neutrophilic inflammation before, but not shortly after (*31*), RSV infection led to significant enhancement of disease (measured by weight loss), with a nadir at 6 to 7 dpi (Fig. 6C). RSV infection resulted in rapid up-regulation of proinflammatory cytokines including IFN-α and IL-6 at 18 hours postinfection (hpi) (Fig. 6D). Production of these cytokines was not significantly altered by rCXCL1 pretreatment. Furthermore, rCXCL1 conditioning did not change the gene expression of *Tnfa* but did coincide with reduced gene expression of *Il1b*, *Cxcl2*, *Cxcl10*, and *Ccl2* in the lung tissue after RSV infection (Fig. 6, E to G). The infiltration of antiviral monocytes (*32*) did not change after rCXCL1 conditioning (Fig. 6H), but neutrophils and neutrophil mediators were increased in the rCXCL1-treated mice compared with mice treated with RSV alone (fig. S8). Nevertheless, rCXCL1 conditioning was associated with a modestly but significantly higher viral load at 18 hpi, which was not sustained to 4 dpi (the peak of viral replication) or 8 dpi (clearance of the virus) (Fig. 6I). Thus, neutrophil infiltration at the time of respiratory viral challenge

transiently alters the inflammatory environment and is associated with a subtle early impairment in the control of virus replication during the asymptomatic period only.

Immunopathology in the murine RSV model causes enhanced disease, which has been shown to be partially mediated by overexuberant T cell activity (*6*, *33*, *34*). In human challenge, $CD8^+$ T cell numbers in bronchial biopsies at 7 dpi showed a nonsignificant positive association with disease severity (P = 0.088; Fig. 7A), suggesting that this may be a contributory mechanism (*20*). To further examine this issue, T cell responses were analyzed in mice pretreated with rCXCL1. At 8 dpi, there was significantly higher total lung cellularity in RSV-infected, rCXCL1-pretreated mice (Fig. 7B) despite the decreasing viral load (Fig. 6I) and resolution of the pulmonary neutrophil response (fig. S9A). Although $CD4^+$ T cell numbers were somewhat increased ($P < 0.05$; Fig. 7C), the most marked elevation in rCXCL1-pretreated mice was in $CD8^+$ T cell numbers ($P < 0.0001$; Fig. 7D), suggesting that these might be the predominant mediators of immunopathology.

Tetramer staining revealed no significant difference in the frequencies of RSV-specific $CD8^+$ T cells in the lung at 8 dpi after rCXCL1 pretreatment (Fig. 7E), but the greater total influx of $CD8^+$ T cells in rCXCL1 pretreated mice resulted in significantly increased numbers of virus-specific $CD8^+$ T cells in the lung around the peak of disease severity (Fig. 7F). Functional markers of these RSV-specific $CD8^+$ T cells were assessed by short-term in vitro restimulation and intracellular cytokine staining (Fig. 7G). Neither granzyme B nor IFN-γ production was enhanced by rCXCL1 pretreatment (Fig. 7H). Depletion of neutrophils by anti-Ly6G antibody treatment in rCXCL1-pretreated mice before RSV inoculation partially abrogated weight loss at 6 and 7 dpi (Fig. 7I). Thus, rCXCL1-driven airway neutrophils at the time of RSV exposure are at least in part responsible for $CD8^+$ T cell recruitment and enhanced disease. Antibody-mediated depletion of $CD8^+$ T cells during the course of infection abrogated weight loss in RSV-infected, rCXCL1-pretreated mice at 6 to 8 dpi (Fig. 7J and fig. S8B). This strongly suggests that $CD8^+$ T cells are the predominant effectors causing enhanced disease. These findings agreed with previous murine studies showing the role of $CD8^+$ T cells in mediating RSV-induced weight loss (*33*).

Preexisting neutrophilic inflammation alters the tissue environment so that the recruitment but not cytotoxic function of $CD8^+$ T cells to the lung is increased later in the disease course. This is associated with transient early changes in soluble mediators and viral load during the asymptomatic period, whereas peak disease is mediated by immunopathology despite the absence of elevated pulmonary neutrophils or viral load at that later time.

Discussion

Using experimental RSV infection of human volunteers and mice, we identified a mechanism of susceptibility and resistance to infection. Before exposure, airway neutrophil activation was associated with enhanced subsequent RSV disease in both mice and human volunteers. This preexisting neutrophilic inflammation was linked to a failure to induce the transient early mucosal inflammatory response to viral exposure dominated by IL-17 and related pathways, which was up-regulated in individuals who resisted infection. Our findings highlight the complex role of the mucosal microenvironment in early responses to respiratory viral infection and demonstrate that the prior state of the mucosa alters the subsequent negotiation between virus and host, which determines the outcome of viral exposure.

Associations between RSV disease and neutrophil- or IL-17–related inflammatory responses during infection have been observed in other settings. For example, IL-17 has been detected in the airway both at the height of severe symptoms and during convalescence in infants with RSV (*35*). However, little is known about its protective role early in the course of infection. IL-17A blockade in clinical trials of secukinumab, an anti–IL-17A monoclonal antibody, caused increased susceptibility to upper respiratory tract infections (*36*–*38*). Furthermore, in a BALB/c murine RSV model, BPZE1 (an attenuated *Bordetella pertussis* vaccine candidate) induced IL-17 and protected against RSV infection (*39*, *40*). Moreover, the intranasal administration of IL-17A 2 hours after RSV challenge decreased disease severity (*41*). These studies support our finding that an IL-17–dominated response during the early phase of infection may be critical for protection before viral replication is established.

The pleiotropic functions of IL-17 are integral to mucosal immunity and include maintenance of barrier integrity, induction of antimicrobial proteins by epithelial cells, and recruitment of inflammatory myeloid cells (*27*). However, here in our controlled studies of adults during the earliest stages after virus exposure, IL-17 signaling was associated with abortive infection. The apparently "silent" incubation phase is therefore immunologically dynamic, representing a critical interval during which clinical outcome is determined. Although an early inflammatory response prevents the establishment of viral infection, subjects destined for symptomatic colds show generalized down-regulation of mucosal defense as the virus gains control. In our studies, preexisting neutrophilic inflammation appeared to prevent this early protective response and was instead associated with enhanced susceptibility to infection and disease. Although we were unable to show an association between the nasal microbiome and susceptibility to RSV infection in our volunteers, our studies do not preclude subtle influences of fungal or viral commensals, which may alter the state of the mucosa, or the effects of recent or subclinical bacterial or viral infections.

Our study has limitations. To avoid traumatic injury to nasal tissue before RSV inoculation, sampling for cytometric analysis (e.g., nasal biopsy) was not performed. The transcriptional and proteomic analyses that we report do not distinguish between elevated neutrophil abundance and greater activity of a static number of neutrophils, nor could we identify the cellular sources of cytokines such as IL-17A that we found to be elevated at baseline and during the presymptomatic viral incubation period. In future studies, we hope to resolve these issues by single-cell analysis of airway cells in the early stages of viral infection. Although our murine studies were unable to replicate all aspects of human respiratory viral infection, they permitted the examination of possible mechanisms by detailed cytometric analysis of the airway and selective recruitment of airway neutrophils through CXCL1 instillation. Enhancement of viral load at 18 hours in mice pretreated with CXCL1 showed that the induction of neutrophilic inflammation in the airway did indeed enhance RSV susceptibility, with an early impairment of antiviral immune responses translating into enhanced disease driven by $CD8^+$ T cells.

Previous studies have demonstrated that activation of one arm of the immune response may simultaneously suppress or disable another. For example, the activation of the T helper (T_H1) transcription factor T-bet limits T_H2 responses by inhibiting GATA3 (*42*). Patients with chronic neutrophilic lung inflammation (such as bronchiectasis, chronic obstructive pulmonary disease, and some types of asthma) are known to have enhanced susceptibility to respiratory virus-induced disease (*43*). Similarly, in naturally acquired human influenza infection, we recently showed that inflammatory signatures characteristic of antibacterial immunity correlated with severe disease, whereas antiviral responses were associated with milder illness (*44*). In murine models of colitis, RIG-I–mediated IL-15 secretion induced by commensal viruses supported the survival of intraepithelial lymphocytes, which could in turn inhibit aberrant inflammation (*45*). Additionally, viral bacteriophages within *Pseudomonas aeruginosa* have been shown to induce TLR3- and TRIF-mediated type I IFN production with associated suppression of phagocytic responses (*46*). These studies thus show that antiviral immunity can suppress responses typically associated with antibacterial and inflammatory

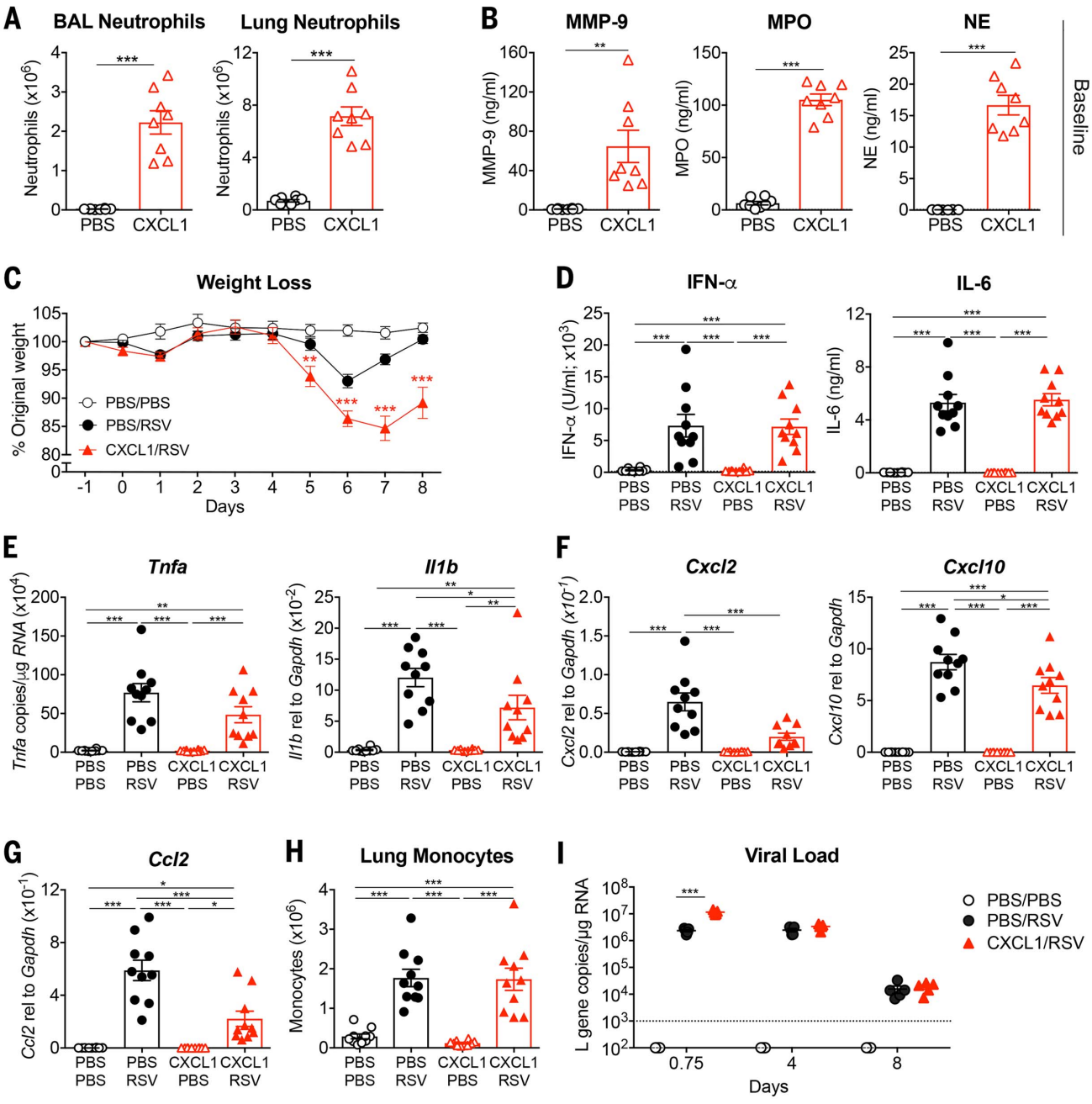


Fig. 6. Mice treated with the neutrophil chemoattractant CXCL1 before RSV infection show an early increase in viral load and develop more severe disease. Mice were treated with mock (PBS) or 10 μg of CXCL1 intranasally. (**A**) At 12 hpi, BAL and lung neutrophils were quantified by flow cytometry and (**B**) MMP-9, MPO, and NE levels in the BAL were quantified by ELISA. Mice were treated with mock (PBS) or 5 to 10 μg of CXCL1 intranasally and after 9 to 12 hours were infected with mock (PBS) or 7.5×10^5 FFUs of RSV intranasally. (**C**) Weight loss as a percentage of original weight. (**D**) IFN-α and IL-6 quantified in the BAL using ELISA. Levels of (**E**) *Il1b* and *Tnfa*, (**F**) *Cxcl2* and *Cxcl10*, and (**G**) *Ccl2* were determined in lung tissue using qPCR. (**H**) Total number of lung monocytes quantified by flow cytometry as previously described (*32*). (**I**) RSV L gene copy numbers were quantified at 0.75, 4, and 8 dpi in lung tissue by qPCR. Data in (A) and (B) are shown as the means ± SEM of eight individual mice per group pooled from two independent experiments. Data in (C) are shown as the means ± SEM of 10 (PBS/PBS) or 14 to 16 (RSV) individual mice pooled from two (PBS/PBS) or three (RSV) independent experiments. Data in (D) to (H) are shown as the means ± SEM of 10 individual mice per group pooled from two independent experiments. Data in (I) are shown as the means ± SEM of five individual mice per group representative of at least two independent experiments, with the dotted line representing the limit of detection of the assay. The statistical significance of differences in (A) and (B) was analyzed using unpaired, two-tailed Student's *t* test. The data in (C) and (I) were analyzed using two-way ANOVA with Bonferroni's post hoc test, and only the statistically significant differences between RSV-infected groups are shown. The data in (D) to (H) were analyzed using one-way ANOVA with Tukey's post hoc test. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

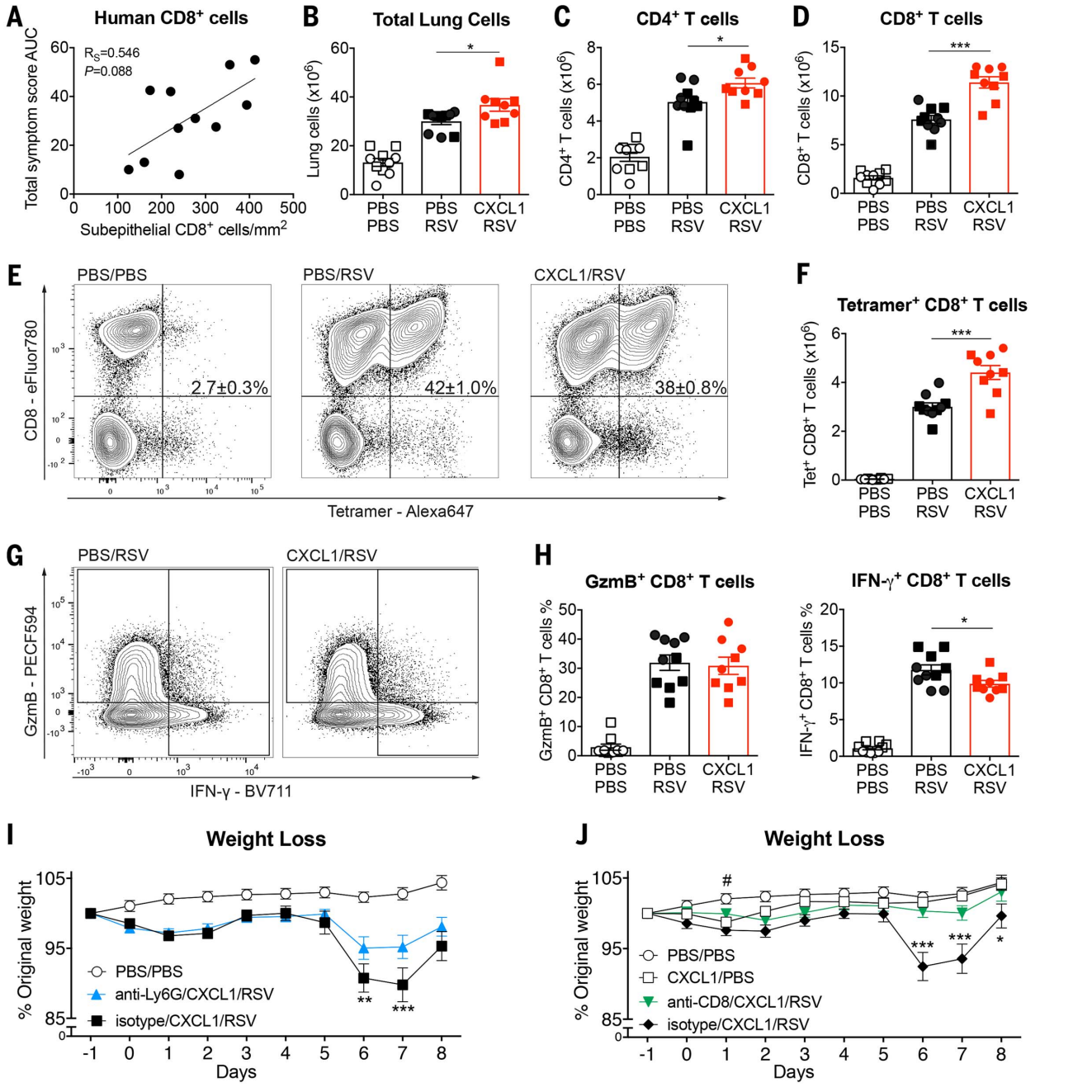


Fig. 7. Mice treated with the neutrophil chemoattractant CXCL1 before RSV infection develop CD8$^+$ T cell–driven disease. (**A**) Correlation between CD8$^+$ T cell numbers in the subepithelial layer of endobronchial biopsy samples from human participants infected with RSV at 7 dpi and cumulative symptom scores over the course of infection (n = 11). (**B** to **H**) Mice were treated with mock (PBS) or 10 μg of CXCL1 intranasally 12 hours before RSV infection. Lungs were analyzed at 8 dpi. (B) Total number of lung cells, (C) lung CD4$^+$ T cells, and (D) lung CD8$^+$ T cells assessed by flow cytometry. (E) Representative flow cytometry plots of lung RSV tetramer–positive CD8$^+$ T cells; the mean percentage ± SEM is indicated in the upper right quadrant. (F) Total number of lung RSV tetramer–positive CD8$^+$ T cells as quantified by flow cytometry. (G) Representative flow cytometry plots of lung granzyme B–positive (GzmB$^+$) and IFN-γ$^+$ CD8$^+$ T cells. (H) Frequency of GzmB$^+$ and IFN-γ$^+$ CD8$^+$ T cells quantified by flow cytometry. (**I**) Anti-Ly6G or isotype control–treated mice were administered PBS or 10 μg of CXCL1 intranasally on day 0 and at 12 hours after being infected with mock (PBS) or 6 to 7.5 × 10^5 FFUs of RSV intranasally. Weight loss is shown as the percentage of original weight. (**J**) Mice treated with anti-CD8 or isotype control antibodies were administered PBS or 10 μg of CXCL1 intranasally on day 0 and at 12 hours after being infected with mock (PBS) or 6 to 7.5 × 10^5 FFUs of RSV intranasally. Weight loss is shown as the percentage of original weight. Data in (B) to (D), (F), and (H) are shown as the means ± SEM of 10 (PBS/PBS, PBS/RSV) or nine (CXCL1/RSV) mice pooled from two independent experiments; repeat 1, 10 μg of CXCL1 (circles) and repeat 2, 8 μg of CXCL1 (squares). Statistical significance of differences in (A) were analyzed using Spearman's rank correlation test, where the R score is given as R_S. Data in (B) to (D), (F), and (H) were analyzed using one-way ANOVA with Tukey's post hoc test. Data in (I) and (J) were analyzed using two-way ANOVA with Bonferroni's post hoc test. Asterisks indicate statistically significant differences between RSV-infected groups. Hash symbol indicates statistically significant differences between PBS/PBS and CXCL1/PBS groups. *#P < 0.05, **P < 0.01, ***P < 0.001.

mechanisms at multiple levels. Together with our present findings, they suggest that there may be opposing mechanisms whereby neutrophilic inflammation before viral exposure antagonizes optimal antiviral defense and results in failure to control viral infection.

We postulate that this apparent competition between alternative types of immune responses may be relevant in a number of infections with protracted incubation periods, including coronaviruses (*47*). The prolonged presymptomatic period that we observed during RSV infection is also evident in SARS-CoV-2 infection, in which the interval between exposure and symptom development is typically ~5.2 days (*48*), compared with 4 to 5 days for RSV shown here. By contrast, during infections such as influenza A, in which there is a very short incubation period [typically ~2 days (*47*)], intrinsic and innate responses may be rapidly overtaken by viral replication. Again, antibody response to SARS-CoV-2 (*49*) and other human coronaviruses (*50*) are apparently short-lived, a feature shared with RSV infection (*3*).

The distinct separation of pulmonary neutrophil and $CD8^+$ T cell infiltration, which show no temporal overlap, implies an indirect relationship between these immune effectors. The early production of IL-17 may be achieved by a variety of cell types, including $CD4^+$ Th17 and $CD8^+$ T cytotoxic 17 T cells, γδ T cells, neutrophils, and type 3 innate lymphoid cells, which are abundant at mucosal surfaces (*51*–*53*). Several mechanisms may link neutrophilic inflammation with enhanced $CD8^+$ T cell recruitment. The preinfection pattern of neutrophil activation is associated with the up-regulation of peptidases and hydrolases, which have been shown to modulate cytokine activity (*54*) in other systems, potentially altering the immediate antiviral response and signaling to natural killer and T cells (*55*). This is supported by the observation that *Il1b*, *Cxcl2*, *Cxcl10*, and *Ccl2* expression, all of which have been implicated in antiviral defense, are perturbed, either directly or through negative feedback. Enhanced $CD8^+$ T cell migration may also result from neutrophil-derived CXCL12-containing trails, as shown in the influenza-infected mouse trachea (*56*). Furthermore, because TLR4 antagonists reduce influenza-induced mortality in mice (*57*), neutrophil-derived oxidized phospholipids (which are TLR4 agonists) may enhance susceptibility. Neutrophil proteases may also degrade antiviral peptides such as cathelicidin (LL-37) (*4*), which would influence immediate susceptibility and might contribute to the early transient difference seen in viral load. The direct enhancement of antigen processing or presentation of antigen by neutrophils to T cells (*58*) is less likely due to the low number of T cells present early after infection, when neutrophils are high, and vice versa later. However, it is possible that neutrophils influence antigen presentation by dendritic cells, synergizing with the early increase in viral load and promoting $CD8^+$ T cell activation, differentiation, and proliferation to enhance viral clearance. The large number of potential contributors means that the precise mechanism(s) through which neutrophils enhance $CD8^+$ T cell recruitment and therefore exacerbated RSV disease remains a major topic for future studies.

Together, our results reveal an integrated, time-dependent mucosal response that critically determines the outcome of pathogen exposure. The susceptibility to infection associated with neutrophilic inflammation before viral exposure reflects a multilayered protective network that reacts rapidly to viral encounter, determining outcome before the presentation of disease. These findings not only increase our capacity to predict previously inexplicable variations in the outcome of exposure to viral pathogens, but also suggest that manipulation of neutrophil-, peptidase-, and IL-17–related pathways may affect antiviral protection, offering additional strategies to prevent disease caused by existing and emerging viral pathogens.

Materials and Methods

Human volunteer infection study design

Healthy, nonsmoking adults 18 to 55 years of age were eligible as previously described except no prescreening for serum-neutralizing antibodies was performed (*16*). As an exploratory study, no power calculations were performed to determine sample size before the study began. Participants provided written informed consent and the study was approved by the UK National Research Ethics Service (study reference nos. 10/H0711/94 and 11/LO/1826). The study was registered on ClinicalTrials.gov under ID no. NCT01349543. Volunteers provided a full clinical history and underwent a physical examination and a panel of screening blood tests to exclude underlying hematological, biochemical, or immune abnormalities. All participants were screened for HIV, hepatitis B virus, and hepatitis C virus. Participants were excluded if they developed common cold symptoms or any acute illness within 6 weeks before inoculation. On the day before inoculation, a nasal lavage specimen was tested using multiplex PCR for a panel of 13 respiratory viruses and bacteria to exclude concurrent upper respiratory tract infection. Volunteers entered quarantine on study day −1, were inoculated by intranasal drops with 10^4 plaque-forming units of RSV strain M37 (Meridian Lifesciences, Memphis, TN, USA; GenBank sequence accession no. KM360090) the next day, and remained in quarantine as described previously (*3*). Volunteers provided daily assessment of their clinical symptoms in the following categories: nasal discharge, nasal obstruction, sneezing, sore throat, headache, fever, and malaise. Each category was scored by volunteers as 0 (absent), 2 (moderate), or 3 (severe). A symptomatic cold was defined as fulfilment of two of the three following conditions: (1) a cumulative symptom score ≥14, (2) nasal discharge lasting ≥3 days, and (3) a subjective assessment of having a cold.

Samples of the nasal mucosal lining fluid were collected daily using precut strips of Leukosorb (Pall Life Sciences, Portsmouth, UK), eluted in 300 μl of PBS, 1% bovine serum albumin (BSA) (w/v), 0.5% Triton X-100 (v/v), and 0.05% sodium azide (w/v) (Sigma-Aldrich, UK), and stored at −80°C for later analysis, as previously described (*59*). Nasal washes were performed as previously described (*3*). Nasal tissue samples were collected at 0 and 3 dpi using ASI-Rhinopro nasal curettage devices (Arlington Scientific, Springville, UT, USA), immediately placed in 800 μl of TRIZol buffer (Life Technologies, Paisley, UK), vortexed thoroughly, and stored at −80°C for later analysis.

Quantification of viral load

RSV load in humans was determined from daily nasal wash samples by qPCR as previously described (*20*).

Immunoassays

Levels of CCL3 (MIP-1α), CCL5 (RANTES), CXCL8 (IL-8), CXCL10 (IP-10), IFN-α, IFN-β, IFN-γ, IL-1β, IL-6, IL-10, IL-15, IL-17A, and TNF-α were assayed using MSD multiplex immunoassays (Mesoscale Diagnostics, Rockville, MD, USA).

RNA-seq of nasal curettage tissue samples

Nasal curettage samples were collected as previously described (*60*) and preserved in TRIzol. After collection, 0.2× chloroform was added to the samples for every 1× of TRIzol, mixed by inverting the tube 10 times, incubated for 5 min, and centrifuged for 10 min at 3500*g*. The supernatant was then transferred into a new tube and total RNA was extracted with an SV96 kit as per the manufacturer's protocol (Promega, Madison, WI, USA).

cDNA libraries were prepared with the TruSeq Stranded Total RNA Library Prep Kit with the Ribo-Zero Human Sample Prep Kit (Illumina, San Diego, CA, USA). Final cDNA libraries were analyzed for size distribution using the Tapestation 2200 (Agilent Technologies), quantitated by qPCR (KAPA Library Quant Kit, KAPA Biosystems, Wilmington, MA, USA), normalized to 2 nM, and sequenced with 2 × 50-bp paired-end reads on a HiSeq 2500 (Illumina) with a minimum of 3 GB of raw RNA-seq data.

RNA-seq reads were mapped to human B38 genome with Omicsoft OSA, and read counts mapped to the ensemble genes (Release 75) were calculated with RSEM. Gene count data

were normalized by DESeq2 (R package, version 1.18.1). An arbitrary filter to remove genes with low counts was applied to the normalized count data (only protein-coding genes and normalized count >35 in at least six samples were kept, which resulted in 14,670 ensembl genes). Wald tests were performed using DESeq2 to identify differential expression of these genes. DEGs were determined as those with $P < 0.01$ and a >0.5 $\log^2$ fold change relative to baseline for the Cold or No Cold group and $P_{adj} < 0.05$ for the Cold versus the No Cold group at baseline. Log^2 normalized count data were used in the heatmap. Gene set overrepresentation analysis of the GO and KEGG pathways was performed with clusterProfiler (R package, version 3.6.0). A coexpression gene network was built with WGCNA (R package, version 1.63) and visualized with Cytoscape (*61*).

Bronchial biopsy immunohistochemistry

Endobronchial biopsies were fixed immediately in 4% paraformaldehyde and paraffin embedded. RSV was stained using a mouse anti-RSV antibody cocktail (NCL-RSV3, Leica Biosystems, UK) at 1:50 dilution. $CD8^+$ T cells were identified by staining with mouse anti-CD8 (M0707, Dako, Denmark) at 1:100 dilution using the EnVision peroxidase-staining method (Dako) as previously described (*62*). Briefly, 5-μm sections were stained according to the manufacturer's instructions, and an irrelevant mouse IgG1 kappa antibody (MOPC21) was used as negative control for staining specificity of mouse monoclonal antibodies. RSV-infected A549 cells were used as positive staining controls. Quantification was achieved as previously described with operators blinded to sample timing and infection status (*62*). Briefly, slides were coded to avoid observer bias, and areas of epithelium and subepithelium were assessed using a Leitz Dialux 20 light microscope, an Apple Macintosh computer, and Image 1.5 software. Total epithelial and subepithelial areas of two to three bronchial biopsies were counted for each bronchoscopy. Cell counts were expressed as the number of cut cell profiles with visible nucleus per 1 mm^2 of subepithelium and per 0.1 mm^2 of epithelium. The coefficient of variation for repeat counts of positive cells by a single observer ranged from 5 to 6%.

Respiratory bacterial load and 16S sequencing

DNA extraction was performed on throat swabs using the MPBio FastDNA spin kit for soil as per the manufacturer's instructions. Blank swabs were extracted and sequenced to control for contamination. Bacterial loads were quantified in triplicate by SYBR green qPCR as previously described (*63*).

Sequencing of the 16*S* rRNA gene was performed on the Illumina MiSeq platform using dual barcode fusion primers targeting the V4 region of the 16*S* rRNA gene as previously described (*63*). Extraction controls, PCR-negative, and mock communities were included on sequencing runs. Sequence processing was performed using QIIME, and UCLUST was used to assign taxonomic identification to operational taxonomic units using the Silva 115 NR database, as previously described (*63*). The operational taxonomic unit and tree files were output into the R package version 3.3.2.

Mouse infection studies

Female wild-type C57BL/6 mice (8 weeks of age) were purchased from Charles River Laboratories. All mice were housed in specific-pathogen-free conditions. All animal experiments were reviewed and approved by the Animal Welfare and Ethical Review Board within Imperial College London and approved by the UK Home Office in accordance with the Animals (Scientific Procedures) Act of 1986 and ARRIVE guidelines.

Plaque-purified human RSV (originally the A2 strain from ATCC, USA) was grown in HEp2 cells (from ATCC, CCL-23) (*64*). Mice were lightly anesthetized before intranasal administration of 100 μl containing 6 to 7.5×10^5 focus-forming units (FFUs) of RSV or PBS control. In some instances, this was preceded by an intranasal instillation of 5 to 10 μg of recombinant murine CXCL1 (BioLegend, San Diego, USA) 12 hours before infection; 500 μg of anti-Ly6G (1A8) or isotype control (2A3) intraperitoneally on −1, 1, and 3 dpi; or 250 μg of anti-CD8 (YTS169.4) or isotype control (LTF-2) intraperitoneally on −1, 2, and 5 dpi (all antibodies from BioXCell, West Lebanon, NH, USA). Mice were sacrificed by a fatal dose of pentobarbital injected intraperitoneally.

To enumerate neutrophils in the BAL, 1 ml of PBS supplemented with 0.5 mM EDTA was used to lavage the lungs three times. The BAL was centrifuged at 800*g* for 10 min, and red blood cells were removed by lysis with ACK buffer (0.15 M NH_4Cl, 1.0 mM $KHCO_3$, and 0.1 mM Na_2EDTA). The cell number was determined by Trypan blue (Thermo Scientific) exclusion of dead cells. The cellular composition was determined by spinning 1 to 2×10^5 cells onto a microscope slide (Thermo Scientific) at 1000*g* for 5 min using a Cytospin 4 Cytocentrifuge (Thermo Scientific). Slides were stained with hematoxylin and eosin using the Reastain Quick-Diff kit (Gentaur, Kampenhout, Belgium) according to the manufacturer's instructions. Cells were classified as neutrophils using a microscope, and ≥300 cells were counted.

Mediators in the BAL were measured using enzyme-linked immunosorbent assay. Murine IFN-α and IL-6 were detected as previously described (*32*). Murine MMP-9, MPO, and NE were detected using DuoSet enzyme-linked immunosorbent assay (ELISA) kits (R&D Systems). Data were acquired on a FLUOstar Omega plate reader (BMG Labtech) and analyzed using MARS data analysis (BMG Labtech).

Lung tissue was homogenized using a TissueLyser LT (Qiagen), and total RNA was extracted from homogenized lung tissue supernatant using the RNeasy Mini kit including DNA removal (Qiagen). RNA concentrations were determined by NanoDrop (Thermo Scientific). RNA was converted to cDNA using the High-Capacity RNA-to-cDNA kit (Applied Biosystems). All primers and probes used are listed in supplementary materials, and all assays used QuantiTect Probe PCR Master Mix (Qiagen). To quantify RSV L gene and *Tnfa* levels in lung tissue (*64*, *65*), the exact number of copies was calculated using a plasmid standard and the results were normalized to the housekeeping gene *Gapdh* (Applied Biosystems). For relative quantification, the expression of *Ccl2*, *Cxcl2*, *Cxcl10*, and *Il1b* (all from Applied Biosystems) were expressed relative to the expression of *Gapdh*. First, the ΔCt (where Ct = cycle threshold) between the target gene and the *Gapdh* for each sample was calculated. Then, the expression was calculated as $2^{-\Delta Ct}$. Analysis was performed using the 7500 Fast System SDS Software (Applied Biosystems).

Flow cytometry and restimulation experiments

For flow cytometry staining, 2.5×10^6 lung cells were incubated for 20 min at 4°C with purified rat IgG2b anti-mouse CD16/CD32 receptor antibody in PBS containing 1% BSA and 5 mM EDTA and then stained with fluorochrome-conjugated antibodies (see the supplementary materials) in PBS for 30 min at 4°C. Cells were fixed with fixation buffer (BioLegend). For RSV M tetramer staining, Alexa Fluor 647–conjugated $M_{187-195}$ tetramers ($H\text{-}2D^b$/NAITNAKII) were obtained from the National Institutes of Health (NIH) Tetramer Core Facility (Emory University, Atlanta, GA, USA). Tetramer staining was performed after Fc block and before surface staining for 30 min at room temperature in PBS containing 1% BSA and 5 mM EDTA. For intracellular staining, the cells were restimulated with 5 μg/ml RSV $M_{187-195}$ peptide (A&A Labs) for 1 hour at 37°C. After the addition of Golgi Plug (BD Biosciences), the samples were incubated for another 3 hours, stained for surface marker expression as described above, and fixed in fixation buffer (BioLegend). Cells were stained with fluorochrome-conjugated antibodies against granzyme B and IFN-γ in the presence of purified rat IgG2b anti-mouse CD16/CD32 receptor antibody in permeabilization buffer (BioLegend) for 1 hour. Analysis was performed on an LSR Fortessa flow cytometer (BD Biosciences). Acquisition was set to 2.5×10^5 single, live, $CD45^+$ cells. The gating strategies are shown in fig. S6. Monocyte and neutrophil enumeration, shown in Fig. 6G and figs. S7A and S8A, was performed using

gating strategies described previously (*32*). Data were analyzed with FlowJo software (Tree Star, Ashland, OR, USA). Total cell populations were quantified as the whole lung count × % population of $CD45^+$ cells × % $CD45^+$ live cells × proportion of lung tissue sampled.

Statistical analyses

Statistical analyses were performed with Prism version 7 software (GraphPad, La Jolla, CA, USA) and R version 3.3.2 software. Mediator concentrations were determined from a standard curve with a four-parameter logistic model fit as per the product literature. Limits of detection for each mediator were calculated as twice the standard deviation of the assay blank well (background signal) duplicates, and samples yielding a signal below this value were assigned a value of this limit. Daily mediator concentrations were expressed as the $\log^2$ fold change over the baseline (day 0) value for each subject and a least-squares locally estimated scatterplot smoothing (LOESS) model was applied using a polynomial degree of 1 and a span of 0.75. A grid of 100 points across the span of days 1 to 10 was used to produce the prediction curves. Pointwise 95% confidence intervals were derived using a bootstrap percentile method as previously described (*3*) except the derived LOESS fit was used to calculate the predicted response values. One property of this LOESS fitting is that the graphical representation of the line at each time point is influenced by the preceding and subsequent data. As such, any apparent differences between groups at baseline results from the difference between groups at day 1 (the next time point). Univariate logistic regression was performed with disease status (Cold or No Cold) as the dependent variable to identify the mediator response(s) associated with inoculation outcome. For the mouse studies, statistical analysis was performed using Prism version 7 software. Data are presented as the mean ± SEM. As indicated, Student's *t* test was used to compare two groups; one-way analysis of variance (ANOVA) followed by Tukey's post hoc test was used to compare multiple groups. The weight curves were analyzed using a two-way ANOVA followed by Bonferroni's post hoc test. $P < 0.05$ was considered statistically significant for all tests ($^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ in all figures).

REFERENCES AND NOTES

1. J. A. Whitsett, T. Alenghat, Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nat. Immunol.* **16**, 27–35 (2015). doi: 10.1038/ni.3045; pmid: 25521682
2. T. T. Hansel, S. L. Johnston, P. J. Openshaw, Microbes and mucosal immune responses in asthma. *Lancet* **381**, 861–873 (2013). doi: 10.1016/S0140-6736(12)62202-8; pmid: 23428115
3. M. S. Habibi *et al*., Impaired antibody-mediated protection and defective IgA B-cell memory in experimental infection of adults with respiratory syncytial virus. *Am. J. Respir. Crit. Care Med.* **191**, 1040–1049 (2015). doi: 10.1164/rccm.201412-2256OC; pmid: 25730467
4. S. M. Currie *et al*., Cathelicidins have direct antiviral activity against respiratory syncytial virus in vitro and protective function in vivo in mice and humans. *J. Immunol.* **196**, 2699–2710 (2016). doi: 10.4049/jimmunol.1502478; pmid: 26873992
5. M. Zanin, P. Baviskar, R. Webster, R. Webby, The interaction between respiratory pathogens and mucus. *Cell Host Microbe* **19**, 159–168 (2016). doi: 10.1016/j.chom.2016.01.001; pmid: 26867175
6. P. J. M. Openshaw, C. Chiu, F. J. Culley, C. Johansson, Protective and harmful immunity to respiratory syncytial virus infection. *Annu. Rev. Immunol.* **35**, 501–532 (2017). doi: 10.1146/annurev-immunol-051116-052206; pmid: 28226227
7. R. J. Geerdink, J. Pillay, L. Meyaard, L. Bont, Neutrophils in respiratory syncytial virus infection: A target for asthma prevention. *J. Allergy Clin. Immunol.* **136**, 838–847 (2015). doi: 10.1016/j.jaci.2015.06.034; pmid: 26277597
8. T. Shi *et al*., Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: A systematic review and modelling study. *Lancet* **390**, 946–958 (2017). doi: 10.1016/S0140-6736(17)30938-8; pmid: 28689664
9. A. R. Falsey, P. A. Hennessey, M. A. Formica, C. Cox, E. E. Walsh, Respiratory syncytial virus infection in elderly and high-risk adults. *N. Engl. J. Med.* **352**, 1749–1759 (2005). doi: 10.1056/NEJMoa043951; pmid: 15858184
10. T. Shi *et al*., Global disease burden estimates of respiratory syncytial virus-associated acute respiratory infection in older adults in 2015: A systematic review and meta-analysis. *J. Infect. Dis.* jiz059 (2019). doi: 10.1093/infdis/jiz059; pmid: 30880339
11. L. Zhang, M. E. Peeples, R. C. Boucher, P. L. Collins, R. J. Pickles, Respiratory syncytial virus infection of human airway epithelial cells is polarized, specific to ciliated cells, and without obvious cytopathology. *J. Virol.* **76**, 5654–5666 (2002). doi: 10.1128/JVI.76.11.5654-5666.2002; pmid: 11991994
12. W. Aherne, T. Bird, S. D. Court, P. S. Gardner, J. McQuillin, Pathological changes in virus infections of the lower respiratory tract in children. *J. Clin. Pathol.* **23**, 7–18 (1970). doi: 10.1136/jcp.23.1.7; pmid: 4909103
13. Committee on Infectious Diseases, From the American Academy of Pediatrics: Policy statements—Modified recommendations for use of palivizumab for prevention of respiratory syncytial virus infections. *Pediatrics* **124**, 1694–1701 (2009). doi: 10.1542/peds.2009-2345; pmid: 19736258
14. J. Vekemans *et al*., Respiratory syncytial virus vaccine research and development: World Health Organization technological roadmap and preferred product characteristics. *Vaccine* **37**, 7394–7395 (2019). doi: 10.1016/j.vaccine.2017.09.092; pmid: 29395536
15. M. C. Crank *et al*., A proof of concept for structure-based vaccine design targeting RSV in humans. *Science* **365**, 505–509 (2019). doi: 10.1126/science.aav9033; pmid: 31371616
16. J. P. DeVincenzo *et al*., Viral load drives disease in humans experimentally infected with respiratory syncytial virus. *Am. J. Respir. Crit. Care Med.* **182**, 1305–1314 (2010). doi: 10.1164/rccm.201002-0221OC; pmid: 20622030
17. A. J. Pollard *et al*., Human microbial challenge: The ultimate animal model. *Lancet Infect. Dis.* **12**, 903–905 (2012). doi: 10.1016/S1473-3099(12)70292-X; pmid: 23174372
18. P. K. Kiyuka *et al*., Human coronavirus NL63 molecular epidemiology and evolutionary patterns in rural coastal Kenya. *J. Infect. Dis.* **217**, 1728–1739 (2018). doi: 10.1093/infdis/jiy098; pmid: 29741740
19. B. Bagga *et al*., Effect of preexisting serum and mucosal antibody on experimental respiratory syncytial virus (RSV) challenge and infection of adults. *J. Infect. Dis.* **212**, 1719–1725 (2015). doi: 10.1093/infdis/jiv281; pmid: 25977264
20. A. Jozwik *et al*., RSV-specific airway resident memory CD8+ T cells and differential disease severity after experimental human infection. *Nat. Commun.* **6**, 10224 (2015). doi: 10.1038/ncomms10224; pmid: 26687547
21. J. P. Devincenzo, Natural infection of infants with respiratory syncytial virus subgroups A and B: A study of frequency, disease severity, and viral load. *Pediatr. Res.* **56**, 914–917 (2004). doi: 10.1203/01.PDR.0000145255.86117.6A; pmid: 15470202
22. C. M. El Saleeby, A. J. Bush, L. M. Harrison, J. A. Aitken, J. P. Devincenzo, Respiratory syncytial virus load, viral dynamics, and disease severity in previously healthy naturally infected children. *J. Infect. Dis.* **204**, 996–1002 (2011). doi: 10.1093/infdis/jir494; pmid: 21881113
23. R. S. Thwaites *et al*., Reduced nasal viral load and IFN responses in infants with respiratory syncytial virus bronchiolitis and respiratory failure. *Am. J. Respir. Crit. Care Med.* **198**, 1074–1084 (2018). doi: 10.1164/rccm.201712-2567OC; pmid: 29688024
24. G. G. Jackson, H. F. Dowling, I. G. Spiesman, A. V. Boand, Transmission of the common cold to volunteers under controlled conditions. I. The common cold as a clinical entity. *AMA Arch. Intern. Med.* **101**, 267–278 (1958). doi: 10.1001/archinte.1958.00260140099015; pmid: 13497324
25. A. J. Stoppelenburg *et al*., Local IL-17A potentiates early neutrophil recruitment to the respiratory tract during severe RSV infection. *PLOS ONE* **8**, e78461 (2013). doi: 10.1371/journal.pone.0078461; pmid: 24194936
26. N. Amatya, A. V. Garg, S. L. Gaffen, IL-17 signaling: The yin and the yang. *Trends Immunol.* **38**, 310–322 (2017). doi: 10.1016/j.it.2017.01.006; pmid: 28254169
27. S. J. Aujla *et al*., IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nat. Med.* **14**, 275–281 (2008). doi: 10.1038/nm1710; pmid: 18264110
28. F. G. Hayden *et al*., Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J. Clin. Invest.* **101**, 643–649 (1998). doi: 10.1172/JCI1355; pmid: 9449698
29. E. G. Nicholson *et al*., Robust cytokine and chemokine response in nasopharyngeal secretions: Association with decreased severity in children with physician diagnosed bronchiolitis. *J. Infect. Dis.* **214**, 649–655 (2016). doi: 10.1093/infdis/jiw191; pmid: 27190183
30. W. A. de Steenhuijsen Piters *et al*., Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am. J. Respir. Crit. Care Med.* **194**, 1104–1115 (2016). doi: 10.1164/rccm.201602-0220OC; pmid: 27135599
31. F. Kirsebom, C. Michalaki, M. Agueda-Oyarzabal, C. Johansson, Neutrophils do not impact viral load or the peak of disease severity during RSV infection. *Sci. Rep.* **10**, 1110 (2020). doi: 10.1038/s41598-020-57969-w; pmid: 31980667
32. M. Goritzka *et al*., Alveolar macrophage-derived type I interferons orchestrate innate immunity to RSV through recruitment of antiviral monocytes. *J. Exp. Med.* **212**, 699–714 (2015). doi: 10.1084/jem.20140825; pmid: 25897172
33. M. J. Cannon, P. J. Openshaw, B. A. Askonas, Cytotoxic T cells clear virus but augment lung pathology in mice infected with respiratory syncytial virus. *J. Exp. Med.* **168**, 1163–1168 (1988). doi: 10.1084/jem.168.3.1163; pmid: 3262705
34. F. C. M. Kirsebom, F. Kausar, R. Nuriev, S. Makris, C. Johansson, Neutrophil recruitment and activation are differentially dependent on MyD88/TRIF and MAVS signaling during RSV infection. *Mucosal Immunol.* **12**, 1244–1255 (2019). doi: 10.1038/s41385-019-0190-0; pmid: 31358860
35. T. E. Faber, H. Groen, M. Welfing, K. J. Jansen, L. J. Bont, Specific increase in local IL-17 production during recovery from primary RSV bronchiolitis. *J. Med. Virol.* **84**, 1084–1088 (2012). doi: 10.1002/jmv.23291; pmid: 22585726
36. R. G. Langley *et al*., Secukinumab in plaque psoriasis—Results of two phase 3 trials. *N. Engl. J. Med.* **371**, 326–338 (2014). doi: 10.1056/NEJMoa1314258; pmid: 25007392
37. P. J. Mease *et al*., Secukinumab inhibition of interleukin-17A in patients with psoriatic arthritis. *N. Engl. J. Med.* **373**, 1329–1339 (2015). doi: 10.1056/NEJMoa1412679; pmid: 26422723
38. I. B. McInnes *et al*., Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): A randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **386**, 1137–1146 (2015). doi: 10.1016/S0140-6736(15)61134-5; pmid: 26135703
39. C. Schnoeller *et al*., Attenuated Bordetella pertussis vaccine protects against respiratory syncytial virus disease via an IL-17-dependent mechanism. *Am. J. Respir. Crit. Care Med.* **189**, 194–202 (2014). pmid: 24261996
40. G. Fedele, M. Bianco, A. S. Debrie, C. Locht, C. M. Ausiello, Attenuated Bordetella pertussis vaccine candidate BPZE1 promotes human dendritic cell CCL21-induced migration and drives a Th1/Th17 response. *J. Immunol.* **186**, 5388–5396 (2011). doi: 10.4049/jimmunol.1003765; pmid: 21430219
41. H. Huang, J. Saravia, D. You, A. J. Shaw, S. A. Cormier, Impaired gamma delta T cell-derived IL-17A and inflammasome activation during early respiratory syncytial virus infection in infants. *Immunol. Cell Biol.* **93**, 126–135 (2015). doi: 10.1038/icb.2014.79; pmid: 25267484
42. J. Zhu *et al*., The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell

program during Th1 cell responses. *Immunity* **37**, 660–673 (2012). doi: 10.1016/j.immuni.2012.09.007; pmid: 23041064
43. A. Jha *et al*., Patterns of systemic and local inflammation in patients with asthma hospitalised with influenza. *Eur. Respir. J.* **54**, 1900949 (2019). doi: 10.1183/13993003.00949-2019; pmid: 31391224
44. J. Dunning *et al*., Progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza. *Nat. Immunol.* **19**, 625–635 (2018). doi: 10.1038/s41590-018-0111-5; pmid: 29777224
45. L. Liu *et al*., Commensal viruses maintain intestinal intraepithelial lymphocytes via noncanonical RIG-I signaling. *Nat. Immunol.* **20**, 1681–1691 (2019). doi: 10.1038/s41590-019-0513-z; pmid: 31636462
46. J. M. Sweere *et al*., Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. *Science* **363**, eaat9691 (2019). doi: 10.1126/science.aat9691; pmid: 30923196
47. X. He *et al*., Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020). doi: 10.1038/s41591-020-0869-5; pmid: 32296168
48. Q. Li *et al*., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020). doi: 10.1056/NEJMoa2001316; pmid: 31995857
49. D. F. Robbiani *et al*., Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals. *bioRxiv* 2020.05.13.092619 (2020). pmid: 32511384
50. K. A. Callow, H. F. Parry, M. Sergeant, D. A. Tyrrell, The time course of the immune response to experimental coronavirus infection of man. *Epidemiol. Infect.* **105**, 435–446 (1990). doi: 10.1017/S0950268800048019; pmid: 2170159
51. C. E. Sutton, L. A. Mielke, K. H. Mills, IL-17-producing γδ T cells and innate lymphoid cells. *Eur. J. Immunol.* **42**, 2221–2231 (2012). doi: 10.1002/eji.201242569; pmid: 22949320
52. H. Hamada *et al*., Tc17, a unique subset of CD8 T cells that can protect against lethal influenza challenge. *J. Immunol.* **182**, 3469–3481 (2009). doi: 10.4049/jimmunol.0801814; pmid: 19265125
53. K. C. De Grove *et al*., Characterization and quantification of innate lymphoid cell subsets in human lung. *PLOS ONE* **11**, e0145961 (2016). doi: 10.1371/journal.pone.0145961; pmid: 26727464
54. D. M. Clancy *et al*., Extracellular neutrophil proteases are efficient regulators of IL-1, IL-33, and IL-36 cytokine activity but poor effectors of microbial killing. *Cell Rep.* **22**, 2937–2950 (2018). doi: 10.1016/j.celrep.2018.02.062; pmid: 29539422
55. J. A. Nicolás-Ávila, J. M. Adrover, A. Hidalgo, Neutrophils in homeostasis, immunity, and cancer. *Immunity* **46**, 15–28 (2017). doi: 10.1016/j.immuni.2016.12.012; pmid: 28099862
56. K. Lim *et al*., Neutrophil trails guide influenza-specific $CD8^+$ T cells in the airways. *Science* **349**, aaa4352 (2015). doi: 10.1126/science.aaa4352; pmid: 26339033
57. K. A. Shirey *et al*., The TLR4 antagonist Eritoran protects mice from lethal influenza infection. *Nature* **497**, 498–502 (2013). doi: 10.1038/nature12118; pmid: 23636320
58. M. M. Hufford *et al*., Influenza-infected neutrophils within the infected lungs act as antigen presenting cells for anti-viral CD8(+) T cells. *PLOS ONE* **7**, e46581 (2012). doi: 10.1371/journal.pone.0046581; pmid: 23056353
59. R. S. Thwaites *et al*., Absorption of nasal and bronchial fluids: Precision sampling of the human respiratory mucosa and laboratory processing of samples. *J. Vis. Exp.* **131**, e56413 (2018). doi: 10.3791/56413; pmid: 29443104
60. J. Dhariwal *et al*., Mucosal type 2 innate lymphoid cells are a key component of the allergic response to aeroallergens. *Am. J. Respir. Crit. Care Med.* **195**, 1586–1596 (2017). doi: 10.1164/rccm.201609-1846OC; pmid: 28085492
61. P. Shannon *et al*., Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi: 10.1101/gr.1239303; pmid: 14597658
62. J. Zhu *et al*., Airway inflammation and illness severity in response to experimental rhinovirus infection in asthma. *Chest* **145**, 1219–1229 (2014). doi: 10.1378/chest.13-1567; pmid: 24457412
63. L. Cuthbertson *et al*., The impact of persistent bacterial bronchitis on the pulmonary microbiome of children. *PLOS ONE* **12**, e0190075 (2017). doi: 10.1371/journal.pone.0190075; pmid: 29281698
64. D. C. Lee *et al*., CD25+ natural regulatory T cells are critical in limiting innate and adaptive immunity and resolving disease following respiratory syncytial virus infection. *J. Virol.* **84**, 8790–8798 (2010). doi: 10.1128/JVI.00796-10 pmid: 20573822
65. M. Goritzka *et al*., Alpha/beta interferon receptor signaling amplifies early proinflammatory cytokine production in the lung during respiratory syncytial virus infection. *J. Virol.* **88**, 6128–6136 (2014). doi: 10.1128/JVI.00333-14; pmid: 24648449

ACKNOWLEDGMENTS

We thank staff at the NIH Tetramer Core Facility at Emory University for providing tetramers and staff at the St. Mary's flow cytometry and animal facilities for their technical support. **Funding:** The authors gratefully acknowledge support from the Wellcome Trust (087805/Z/08/Z, 102126/B/13/Z, and 109058/Z/15/Z); The Medical Research Council HIC-Vac network (MR/R005982/1); RSV Consortium in Europe (RESCEU) Horizon 2020 Framework Grant 116019; the UK National Institute for Health Research (NIHR) Comprehensive Local Research Networks (CLRNs); European Respiratory Society and the Asociación Latinoamericana de Tórax joint long-term Research Fellowship 2019 (LTRF 201901-00546); an NIHR Senior Investigator award; the Biomedical Research Centre (NIHR Imperial BRC); and the Health Protection Research Unit in Respiratory Infections at Imperial College London (NIHR HPRU RI). **Author contributions:** P.J.M.O. and C.C. conceived the study. M.S.H., A.P., and C.C. performed infections and collected volunteer samples. M.S.H., R.S.T., A.J., L.C., P.J., and C.C. collected human data. Murine experiments were designed, performed, and analyzed by F.K., A.V., A.O., and C.J. Human data were analyzed by M.S.H., R.S.T., M.C., L.C., T.T., D.N., T.T.H., M.F.M., C.C., and P.J.M.O. The manuscript was written by M.S.H., R.S.T., C.J., C.C., and P.J.M.O. with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The RNA-seq data discussed have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE155237. Sequences from 16S rRNA sequencing have been deposited in the European Nucleotide Archive with the accession codes PRJEB28323 (ERP110514). All other data needed to evaluate the conclusions in this study are present in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/eaba9301/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S14
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

17 January 2020; resubmitted 30 June 2020
Accepted 19 August 2020
10.1126/science.aba9301

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Conformational states dynamically populated by a kinase determine its function

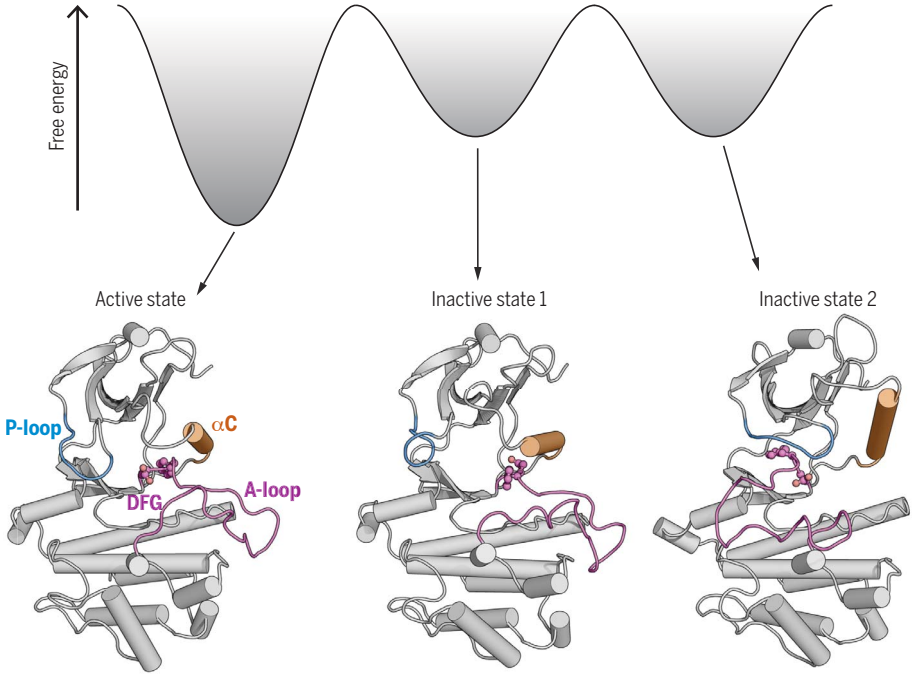
Tao Xie, Tamjeed Saleh, Paolo Rossi, Charalampos G. Kalodimos*

INTRODUCTION: Protein kinases mediate many cell signaling processes. Central to their physiological function is the regulation of their binding and enzymatic activities, which is typically achieved by conformational transitions between active and inactive states. Dysregulation of kinase activity by deletions or mutations often results in disease. Protein kinases are dynamic molecules that intrinsically sample a number of conformational states. However, it has been challenging to experimentally access their conformational ensemble and structurally characterize the discrete conformations associated with distinct activities. Such information could advance our understanding of activation and inhibition mechanisms in this protein family and aid in the development of selective inhibitors.

RATIONALE: We used nuclear magnetic resonance spectroscopy to monitor in atomic-level detail how Abl kinase transitions between distinct conformational states and to elucidate how the conformational ensemble is exploited by mutants, ligands, posttranslational modifications, and inhibitors to regulate the kinase activity and function. We combined structural and energetic approaches to quantitate the contribution of key structural elements such as the activation loop, the Asp-Phe-Gly (DFG) motif, the regulatory spine, and the gatekeeper residue to kinase regulation and provide the mechanistic basis for drug resistance.

RESULTS: We found that the Abl kinase domain interconverts between an active and two, transiently populated, conformational states that adopt discrete inactive structures. There are extensive differences in key structural elements between the conformational states that reveal multiple intrinsic regulatory mechanisms. The small energy difference between active and inactive states allows oncogenic mutations in the regulatory spine or the gatekeeper position to counteract inhibitory mechanisms and constitutively activate the kinase. By capturing and structurally characterizing the conformational state to which the cancer drug imatinib selectively binds, we explain a number of drug-resistance variants isolated in patients. These mutants confer resistance by depleting, through various mechanisms, the conformation to which imatinib binds. To determine the basis for allosteric regulation, we studied a construct that includes the kinase domain and the regulatory domains and that can adopt an assembled and an extended conformation. In the assembled conformation, in which the regulatory domains dock onto the back of the kinase domain, one of the inactive states is selectively stabilized, thereby suppressing catalytic activity. In the extended conformation, wherein the regulatory domains dock on top of the N-lobe, the inactive state is eliminated, thus explaining the increased leukemogenic activity associated with this conformational state. Only one of the detected inactive states appears to be physiologically relevant. The inactive state with no apparent biological function can nevertheless be leveraged for the design of selective inhibitors. Targeting nonphysiological conformational states may be an effective strategy in the design of drugs with increased selectivity and reduced selection pressure for the occurrence of drug-resistance mutations. Although the structure of inactive states can, in principle, vary considerably among kinases, structural comparison of the Abl inactive states with those previously determined for other kinases reveals that there may be a limited number of structurally divergent inactive states intrinsic to kinases.

CONCLUSION: Our data demonstrate that the detection and structural characterization of the distinct conformational states populated by a kinase, coupled to the energetic dissection of the contribution of key structural elements to the selective stability of these states, are essential to advance our understanding of the mechanisms underpinning kinase regulation and function. The approaches presented here can be extended to other kinases to characterize transiently populated conformational states, with the goal of revealing the full repertoire of regulatory and drug-resistance mechanisms in the kinome. ■



Transitions of Abl kinase between conformational states. The Abl kinase domain adopts predominantly (~90%) an active state in solution, but it transiently switches between two low-populated (~5%) states that adopt distinct inactive conformations. Key structural elements that rearrange in the various states are highlighted. The conformational equilibrium is exploited by physiological and pathological stimuli to alter the function of Abl.

The list of author affiliations is available in the full article online.
Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.
*Corresponding author. Email: babis.kalodimos@stjude.org
Cite this article as T. Xie *et al.*, *Science* **370**, eabc2754 (2020). DOI: 10.1126/science.abc2754

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abc2754

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Conformational states dynamically populated by a kinase determine its function

Tao Xie, Tamjeed Saleh, Paolo Rossi, Charalampos G. Kalodimos*

Protein kinases intrinsically sample a number of conformational states with distinct catalytic and binding activities. We used nuclear magnetic resonance spectroscopy to describe in atomic-level detail how Abl kinase interconverts between an active and two discrete inactive structures. Extensive differences in key structural elements between the conformational states give rise to multiple intrinsic regulatory mechanisms. The findings explain how oncogenic mutants can counteract inhibitory mechanisms to constitutively activate the kinase. Energetic dissection revealed the contributions of the activation loop, the Asp-Phe-Gly (DFG) motif, the regulatory spine, and the gatekeeper residue to kinase regulation. Characterization of the transient conformation to which the drug imatinib binds enabled the elucidation of drug-resistance mechanisms. Structural insight into inactive states highlights how they can be leveraged for the design of selective inhibitors.

Protein kinases are key enzymes involved in mediating cell signaling by phosphorylating tyrosine, serine, or threonine residues in target proteins. The activity of protein kinases is tightly regulated, typically by means of intramolecular interactions, binding of protein partners, and post-translational modifications (*1*–*3*). Dysregulation of kinase activity by deletions or mutations often results in cancer, inflammation, or neurodegeneration (*4*, *5*), and disease-associated kinases are the target of therapeutic intervention by small molecules (*6*, *7*). Many genetic mutations associated with cancer affect kinase domains (*8*).

Evidence from biophysical experiments and computational studies, as well as from structures of kinases in complex with various inhibitors, indicates that protein kinases are dynamic and sample alternate conformational states (*9*–*17*). However, it has proved challenging to monitor these transitions and structurally characterize the manifold of conformational states populated by a kinase. Knowledge of the conformational ensemble could give insight into the mechanisms of regulation. Evolutionary pressure has forced kinases to adopt similar active states so that regions required for catalysis are precisely positioned (*18*, *19*). By contrast, inactive states can be distinct and may in principle vary considerably between kinases. Structural information of the inactive states adopted by kinases could aid in the development of selective inhibitors.

Three key structural elements typically determine whether a kinase adopts the active or an inactive state (*18*, *19*): the catalytic Asp-Phe-Gly (DFG) motif, the activation loop (A-loop), and the αC helix (Fig. 1A). In the fully active state, the DFG motif adopts the so-called "DFG-in" state, wherein the Asp (Asp^{400} in Abl) of DFG points inside the active site cleft and is poised to coordinate Mg–adenosine triphosphate (ATP); the A-loop adopts a so-called open conformation that allows substrates to dock onto it to be phosphorylated; and the αC helix adopts the so-called "αC-in" state wherein a conserved αC helix Glu residue (Glu^{305} in Abl) forms an ion pair with a conserved Lys residue (Lys^{290} in Abl) in the β3 strand that coordinates the α and β phosphates of the ATP. A substantial deviation from this architecture will result in an inactive state.

Abl kinase is involved in a number of signaling pathways that control cell growth, survival, invasion, adhesion, and migration (*20*–*22*). The breakpoint cluster region (Bcr)–Abl, an aberrant fusion form of Abl with dysregulated activity, causes leukemia and other cancers and is the target of a number of inhibitors, such as imatinib (Gleevec) (*20*, *23*). However, a majority of patients develop mutations that confer resistance to imatinib and other inhibitors (*24*–*28*). Although it has been possible to explain how some of these mutations decrease the inhibitor affinity, the mechanistic basis for the effect of mutations in sites remote to the inhibitor is not known. We have used nuclear magnetic resonance (NMR) spectroscopy to structurally characterize how Abl switches between the active and two distinct inactive states and to dissect how the conformational ensemble is exploited by mutants, ligands, posttranslational modifications, and inhibitors to regulate the kinase activity and function.

Accessing alternate conformational states of Abl kinase

NMR has provided a wealth of information on the role of dynamics in the regulation of kinases (*14*, *29*–*37*). The ^{1}H-^{15}N– and ^{1}H-^{13}C–correlated NMR spectra of the Abl kinase domain (hereafter Abl; residues 248 to 534) labeled in methyl-bearing (Ala, Ile, Met, Leu, Thr, and Val) and aromatic residues were of high sensitivity and resolution (fig. S1). In the spectra of unliganded Abl recorded at 20°C or higher, the resonances of the A-loop and nearby residues at the αC helix (Fig. 1A) broaden beyond detection (fig. S1, B and C). Such extreme line broadening is commonly observed in kinases studied by NMR in their unliganded state (*11*, *29*) and has typically prevented a detailed characterization of the A-loop. The A-loop encompasses the DFG motif, forms the docking site for substrate binding, and, in tyrosine kinases, includes the tyrosine residue(s) that are phosphorylated to regulate activity (*38*). We obtained near-complete methyl assignment of Abl by lowering the temperature to 10°C, where all of the expected resonances were visible (fig. S1, B and C).

The observed line broadening is due to millisecond–time scale conformational transitions, indicating the presence of alternate conformational states (*39*). To access the conformational states sampled by Abl, we used the chemical exchange saturation transfer (CEST) NMR technique (*40*, *41*). In this experiment, a weak radiofrequency field is applied to monitor magnetization transfer in sites that undergo conformational exchange. We recorded ^{13}C-based CEST experiments on Abl samples labeled (^{1}H, ^{13}C) specifically in the 104 methyl-bearing residues (Fig. 1B and fig. S1A), which afforded high sensitivity even at low temperatures. A CEST profile was obtained for each one of the methyl probes in Abl. If exchange is present (with an exchange rate, k_{ex}, between ~20 and 500 s^{-1}), a large dip at the chemical shift of the dominant, energetically ground state (G) and a smaller dip at the chemical shift of the alternate, energetically excited state are observed (Fig. 1C and fig. S2). Methyl probes reporting on the conformational transitions are distributed throughout the protein, with many of them being located in or near the A-loop and the αC helix (Fig. 1B).

Forty methyl probes (colored cyan in Fig. 1B) showed one minor dip in their CEST profile (e.g., $Val^{318^{CG1}}$ and $Ile^{333^{CD1}}$; Fig. 1C), whereas 43 methyl probes (colored magenta in Fig. 1B) showed two minor dips, indicating the presence of two distinct energetically excited conformational states (e.g., $Leu^{267^{CD2}}$ and $Leu^{389^{CD2}}$; Fig. 1C) referred to as E_1 and E_2. CEST data fitting (see Materials and methods) indicated a linear equilibrium, $G \leftrightarrow E_1 \leftrightarrow E_2$ (Fig. 1D). The exchange rate k_{ex} for the $G \leftrightarrow E_1$ transition is 46.8 ± 4.3 s^{-1} and for the $E_1 \leftrightarrow E_2$ transition is 88.7 ± 13.5 s^{-1}. The population of the ground state (p_G) is 88%, whereas the population of either E_1 (p_{E1}) or E_2 (p_{E2}) is 6%, indicating

Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.
*Corresponding author. Email: babis.kalodimos@stjude.org

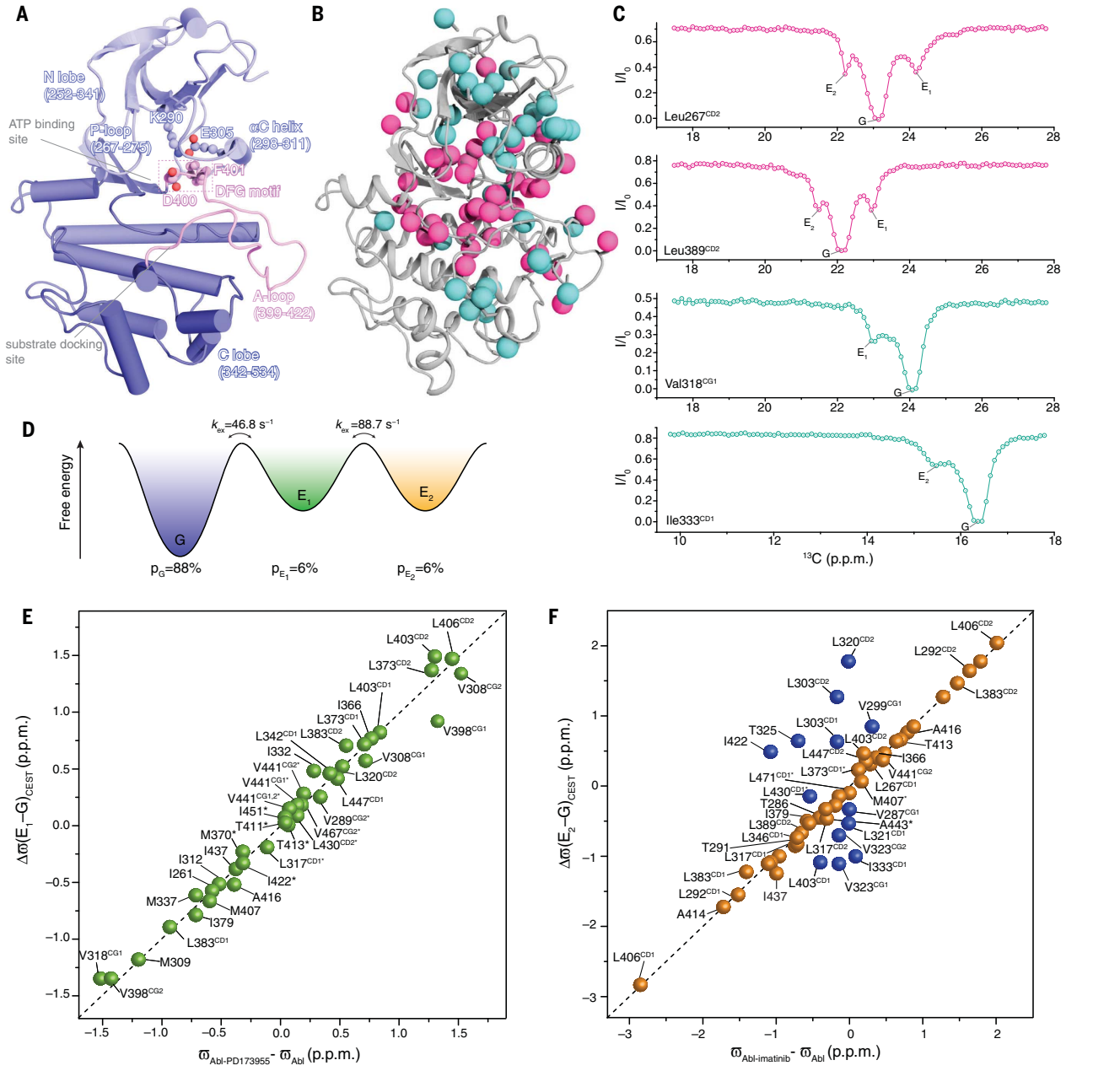


Fig. 1. Characterization of the energetically excited conformational states in Abl kinase by NMR CEST experiments. (**A**) Structure of the Abl kinase domain in the active conformation determined in the current work. Key structural features are highlighted. (**B**) Methyl groups in Abl indicating the presence of one (colored cyan) or two (colored magenta) excited conformational states in the NMR CEST experiments. (**C**) Representative ^{13}C CEST profiles of the indicated methyl groups. The major dip corresponds to the major (ground) state, whereas the minor dips correspond to the energetically excited conformational states E_1 and E_2. As explained in the main text, excited states E_1 and E_2 correspond to inactive states I_1 and I_2, respectively. The ratio I/I_0 is plotted, where I is the intensity after irradiation and I_0 is the intensity without irradiation. (**D**) Energy landscape of the ground (G), E_1, and E_2 states denoting their populations and kinetics of interconversion, as measured by fitting the CEST data. (**E**) Correlation of the CEST-derived chemical shift difference between the ground and E_1 states [$\Delta\varpi(E_1–G)_{CEST}$] with the chemical shift difference between apo (unbound) and PD173955-bound Abl. Methyls (marked with an asterisk) for which the dips corresponding to the ground and E_1 states are within ±0.3 ppm cannot be resolved in the CEST profile, and their chemical shift difference was extracted from ^1H-^{13}C–correlated experiments (see Materials and methods). Methyls that are close (within 6 Å) to PD173955 were excluded from the correlation because their chemical shifts are directly affected by the inhibitor. $Val^{398^{CG1}}$ deviates from linearity because of the rearrangement of the nearby Phe^{401} side chain in the inhibitor complex compared with the E_1 state. (**F**) Correlation of the CEST-derived chemical shift difference between the ground and E_2 states [$\Delta\varpi(E_2–G)_{CEST}$] with the chemical shift difference between apo and imatinib-bound Abl. Methyls that are close (within 6 Å) to imatinib were excluded from the correlation because their chemical shifts are directly affected by the inhibitor. Methyls colored blue deviate from linearity, as discussed in the text and in fig. S8. Methyls (marked with an asterisk) for which the dips corresponding to the ground and E_2 states are within ±0.3 ppm cannot be resolved in the CEST profile, and their chemical shift difference was extracted from ^1H-^{13}C–correlated experiments. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

that the two excited states are equally stable (Fig. 1D).

Structure of the Abl ground state

The structure of unliganded Abl is not known. We used NMR spectroscopy to determine the structure of the ground state (population 88%; Fig. 1D and fig. S1, B and C) of the isolated Abl kinase domain in solution. The structural data demonstrate that the unliganded Abl kinase domain adopts a catalytically active conformation in solution (Fig. 2A and fig. S3, A to C), with the A-loop adopting an open conformation and the DFG motif and αC helix both in the "in" state (Fig. 2A). Phe^{401} of the DFG motif fits snugly into a hydrophobic pocket lined by Val^{308}, Met^{309}, Ile^{312}, Leu^{317}, Val^{318}, and Leu^{373}, and these interactions further stabilize the DFG motif in the "in" state. Tyr^{412}, the residue that when phosphorylated stimulates Abl activity (*42*), forms hydrogen bonds through its side chain to Arg^{381} and Arg^{405}, and these interactions stabilize the open conformation of the A-loop. The phosphate-binding loop (P-loop; also referred to as the Gly-rich loop) is an important region in kinases because it directly affects ligand binding in the active site (*19*, *43*). Tyr^{272}, the bulkiest residue in the P-loop, points inward and occupies the cleft where the phosphate group of ATP is located (Fig. 2A). Thus, Abl adopts predominantly the active state in solution, and inhibitors, such as dasatinib, that selectively target the Abl active state bind with high affinity (fig. S3, D and E) (*44*).

Structure of the Abl excited state 1 (E_1)

To gain structural insight into the low-populated conformational states that Abl samples, we directly measured the ^{13}C chemical shift of the methyls in these two states from the corresponding CEST profiles (Fig. 1C). We next compared the chemical shifts of these two states to the chemical shifts of Abl in complex with 12 different inhibitors (fig. S4). Structural data (*28*, *45*–*52*) have shown that these inhibitors can capture Abl in distinct conformational states, and thus we asked whether any of the low-populated conformational states sampled by unliganded Abl resemble an inhibitor-bound state.

The ^{13}C methyl chemical shifts of the Abl E_1 state showed notable similarity to the chemical shifts of Abl in complex with inhibitor PD173955 (Fig. 1E and fig. S5A). To structurally characterize Abl E_1, we sought to increase its population at the expense of the ground state, as shown before for other systems (*53*–*55*). The double amino acid substitution Met^{309}→Leu/His^{415}→Pro (M309L/H415P) combined with adjusted pH buffer conditions collectively increased the population of E_1 from 6 to 50% (fig. S5, B to D). NMR data showed that the double amino acid substitution M309L/H415P shifts the equilibrium toward E_1 without eliciting other structural changes (fig. S5, B and E). We used the $Abl^{M309L/H415P}$ variant to obtain nuclear Overhauser effect (NOE)–based distance restraints to gain insight into the structure of Abl E_1 (fig. S5, F and G). The structural data showed that Abl in E_1 adopts a structure (Fig. 2B) that is distinct and uncommon among unliganded protein kinases with known structures. Specifically, the DFG motif is flipped 180°, with respect to the active conformation, and it adopts the DFG-out conformation. However, the A-loop remains in an open conformation similar to the active conformation. Because of the DFG flip, the five residues following the DFG motif (Leu^{403} to Met^{407}) of the A-loop are also flipped by 180°. Thus, although the overall structural disposition of the N-terminal part of the A-loop is very similar in the active and in the E_1 states, the backbone dihedral angles are very different, so that the side chains of the residues point in opposite directions (Fig. 2B). Asp^{400} is in the "out" conformation and swaps positions with Phe^{401} of the active state. Thus, Asp^{400} is now located inside a hydrophobic pocket, and, to compensate for this unfavorable environment, its side chain forms a hydrogen bond with the carbonyl oxygen of Val^{318}. With Phe^{401} flipping inside the ATP-binding pocket, Leu^{403} swings in the opposite direction. As a result, Arg^{405}, which in the active conformation forms a hydrogen bond with Tyr^{412} (Fig. 2A), pivots outward and forms a salt bridge with Glu^{311} of the αC helix. The αC helix adopts the "in" conformation where Glu^{305} forms an ion pair with Lys^{290}, similarly to the active state (Fig. 2B). Overall, Abl in E_1 adopts the DFG-out, αC helix–in conformation, and thus it is catalytically inactive. We will refer to E_1 as inactive state 1 (I_1). Structural and NMR data analyses are consistent with Abl in state I_1 adopting a structure that is very similar to the crystal structure of the complex with the PD173955 inhibitor (*46*). Thus, PD173955 binds to and stabilizes a conformational state that already preexists in the conformational ensemble of unliganded Abl (fig. S5H).

Structure of the Abl excited state 2 (E_2)

The ^{13}C methyl chemical shifts of Abl E_2 show an excellent correlation with the chemical shifts of Abl in complex with imatinib (Fig. 1F). The crystal structure of Abl in complex with imatinib showed that a hallmark is the closed conformation of the A-loop (*46*). All of the methyl probes in or around the A-loop experience large chemical shift changes ($\Delta\varpi$) as a result of the A-loop undergoing a large conformational change on imatinib binding. For example, the two methyl groups of Leu^{406} have very characteristic chemical shifts; methyl CD1 has a ^{13}C $\Delta\varpi$ of about −2.8 parts per million (ppm) and methyl CD2 a ^{13}C $\Delta\varpi$ of about 2.1 ppm (Fig. 1F). The excellent correlation of the $\Delta\varpi$ values between E_2 and the imatinib-bound state observed indicates that the A-loop in E_2 adopts the closed conformation. Of note, a number of methyl probes on or nearby the αC helix deviate considerably from linearity (blue spheres in Fig. 1F), suggesting that the αC helix adopts a different conformation in Abl E_2 than that observed in the complex with imatinib.

To gain more detailed structural insight into the Abl E_2 state, we sought to increase its population at the expense of the ground state as we did above for the Abl E_1 state. From the analysis of the crystal structure of the Abl-imatinib complex (*46*), augmented by structural modeling of the unliganded Abl aided by the determined chemical shifts in the E_2 state, we reasoned that the T408Y (T, Thr; Y, Tyr) substitution should stabilize the closed A-loop conformation by forming a hydrogen bond between the hydroxyl of Tyr^{408} and the side chain of Asp^{344} or Asn^{341} (fig. S6A). NMR analysis of the Abl^{T408Y} variant showed that indeed the substitution increases the population of E_2 from 5 to 70% (fig. S6B). Because the exchange between the active and inactive states is slow on the NMR time scale, two sets of peaks are present in the NMR spectra of Abl^{T408Y}: The chemical shifts of one of the sets correspond to those of the ground, active state, whereas the chemical shifts of the second set correspond to the chemical shifts of the E_2 state, as measured by CEST (fig. S6B). The two sets of peaks are connected in a magnetization exchange experiment (fig. S6C), further corroborating that the two states are in dynamic equilibrium. To further increase the population of the E_2 state, we produced the M309L/T408Y double substitution (fig. S6D). The NMR data clearly show that the single T408Y or the double M309L/T408Y substitution shifts the equilibrium toward E_2 without eliciting other structural changes (fig. S6, B and E). The $Abl^{M309L/T408Y}$ and similar variants (Materials and methods), which have E_2 populated at ~90 to 100%, allowed us to measure a large number of NOE-based distance restraints (fig. S7A) that were used to determine the structure of the Abl I_2 state (fig. S7B).

The structure of the Abl E_2 state is depicted in Fig. 2C. Compared with the active state, the A-loop in the E_2 state rotates, with the middle of the A-loop translating by as much as 35 Å (see dashed arrow in Fig. 2C). A segment in the C terminus of the A-loop folds back and masks the substrate docking site while Tyr^{412} occupies the substrate Tyr position, thereby preventing substrate binding altogether. Similar to the Abl I_1 state, the DFG motif in Abl E_2 adopts the "out" conformation. However, Phe^{401} of the DFG in the E_2 state not only flips into the catalytic pocket, as was observed in I_1 (Fig. 2B), but also translates by ~11 Å to occupy a hydrophobic pocket that is lined by Leu^{267}, Val^{275}, Ala^{288}, Phe^{336}, and Leu^{403} (Fig. 2C). This hydrophobic

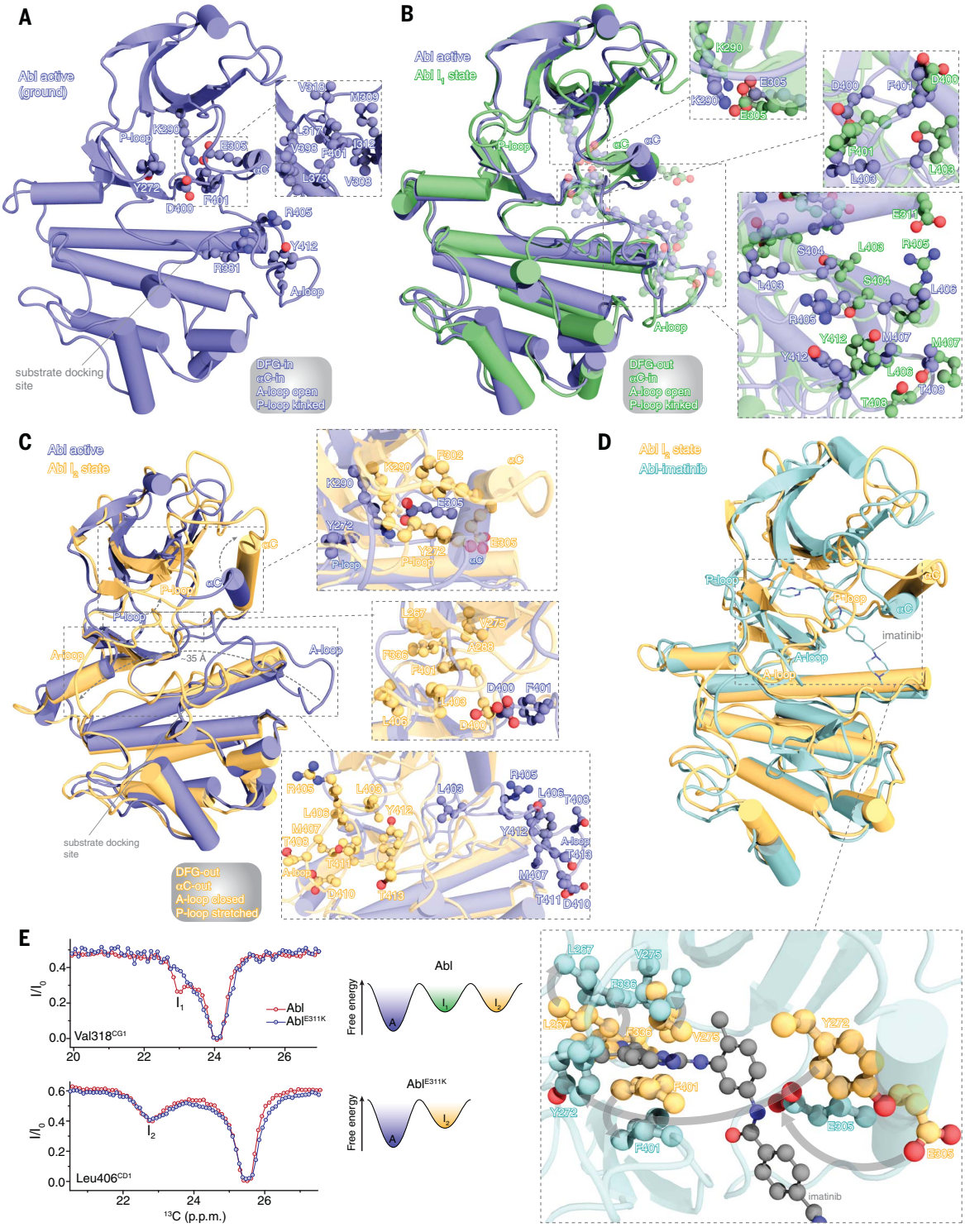


Fig. 2. Structures of the ground (active), E_1 (I_1), and E_2 (I_2) states of Abl. (**A**) NMR solution structure of the ground state of Abl shows that the Abl kinase domain inherently adopts the fully active state. The magnified view shows the hydrophobic residues surrounding Phe^{401}. (**B**) Structure of the Abl E_1 state (green) superimposed on the structure of the active state (blue). Because Abl E_1 adopts an inactive state, it is referred to as I_1. The magnified views highlight similarities (e.g., αC helix) and differences (e.g., DFG motif and A-loop) in the disposition of key structural elements between the two states. (**C**) Structure of the Abl E_2 state (orange) superimposed on the structure of the active state (blue). Because Abl E_2 adopts an inactive state, it is referred to as I_2. The major structural rearrangement undergone by the αC helix, the A-loop, and the P-loop between the two states is indicated by a dashed arrow. (**D**) Superposition of the structures of the Abl I_2 state (orange) and the Abl-imatinib complex [cyan; Protein Data Bank (PDB) ID 1IEP]. The magnified view highlights the structural rearrangement that occurs on imatinib binding in the residues lining the pocket. (**E**) ^{13}C CEST profiles of Abl and Abl^{E311K} demonstrate that the E311K substitution destabilizes the Abl I_1 state but has no apparent effect on the I_2 state. "A" denotes active state. The E311K substitution abrogates the ion pair between Glu^{311} and Arg^{405} in the I_1 state (B).

pocket is formed as a result of the void created by Tyr^{272} (P-loop) swinging outward and is further stabilized and enlarged by the repositioning of A-loop residues Leu^{403} and Leu^{406} (Fig. 2C). The positioning of Phe^{401} into this hydrophobic pocket directly blocks binding of ATP or competitive inhibitors.

In the E_2 state, the entire P-loop undergoes a pronounced rotation and translation with Tyr^{272}, located in the middle of the P-loop, shifting by ~16 Å toward the αC helix (Fig. 2C). The repositioned Tyr^{272} acts as a wedge to push the αC helix away from the C-lobe. As a result, Glu^{305} in the αC helix rotates outward by ~90° and breaks its ion pair with Lys^{290}, an interaction that is required for catalysis (*9*) (Fig. 2C). Thus, the αC helix transitions from the "in" conformation in the active state to the "out" conformation in the E_2 state. Taken together, Abl in the E_2 state adopts a fully inactive conformation with the four key structural elements—A-loop, DFG motif, αC helix, and P-loop—all being in conformations that are not compatible with either substrate binding or ATP binding or hydrolysis. We will refer to the Abl E_2 state as inactive state 2 (I_2).

Structural differences between Abl I_2 and the Abl-imatinib complex

Comparison of the structure of Abl in the I_2 state and the crystal structure of Abl in complex with imatinib (*46*) indicates similarities but also key differences (Fig. 2D). The most important similarity is the conformation of the A-loop that adopts a closed conformation in both the I_2 state and the imatinib complex. Imatinib pushes Phe^{401} toward the C-lobe by ~3 Å to make room for its pyrimidine ring inside the hydrophobic pocket. The DFG motif is in the "out" conformation both in I_2 and the imatinib complex. However, imatinib binding elicits a pronounced change to the structure and disposition of the P-loop and the αC helix. Strong imatinib binding to Abl entails the formation of a hydrogen bond between its secondary amino group and the carboxyl side chain of Glu^{305} (*46*). This interaction forces the αC helix to the "in" conformation. In addition, the P-loop switches from the stretched conformation in I_2 that blocks imatinib from entering into the binding pocket to the kinked conformation observed in the structure of the imatinib complex (Fig. 2D). Thus, imatinib apparently binds selectively to the Abl I_2 state, a low-populated conformational state that resembles the final imatinib complex structure in the A-loop and DFG motif but is very different in the αC helix and P-loop (Fig. 3A). These structural differences explain (fig. S8) the discrepancy between the CEST-derived chemical shifts of certain residues in the Abl I_2 state and their observed chemical shifts in the Abl-imatinib complex (Fig. 1F). Our data suggest that imatinib expends energy to elicit the structural changes required for its binding to Abl, likely resulting in a lower affinity. By contrast, the PD173955 inhibitor binds to the I_1 state with minimal structural changes (fig. S5H), which could explain the higher affinity and potency of PD173955 compared with imatinib for Bcr-Abl, despite the fewer contacts between the inhibitor and Abl in the PD173955 complex (*46*).

Mechanisms underpinning imatinib-resistance mutations

Imatinib is a selective inhibitor of Abl kinase (*56*, *57*) and is widely used to target Bcr-Abl in chronic myelogenous leukemia (CML) patients (*23*). However, a serious problem is the occurrence of mutations that confer resistance to imatinib in a majority of the patients (*24*–*27*, *58*). The underlying mechanism of action is straightforward to understand in mutations—for example, T334I (I, Ile)—that occur in positions that line the drug-binding site and sterically clash with imatinib to decrease its affinity for Abl (Fig. 3B). By contrast, the mechanisms conferring drug resistance by mutations in positions remote to the imatinib site have remained enigmatic.

Substitution of His^{415} by Pro confers resistance to imatinib, even though this residue is located ~18 Å from the inhibitor (Fig. 3B). We find that Abl^{H415P} binds to imatinib with an affinity that is five times weaker than the binding affinity of wild-type Abl for imatinib (Fig. 3C and fig. S9), consistent with a previous report (*59*). To understand the basis for the affinity decrease, we used NMR CEST experiments to measure the effect of the substitution on the Abl conformational ensemble. The data show that H415P destabilizes the Abl I_2 state to the extent that it is not detectable by the CEST experiments, indicating that the population of I_2 in Abl^{H415P} is less than 1% (Fig. 3D). Depletion of the state to which imatinib selectively binds (Fig. 2D) will result in a decrease in imatinib affinity for Abl (Fig. 3C). The destabilization of the I_2 state by the H415P substitution results in the stabilization of the I_1 state, with its population increasing twofold (Fig. 3D). The structural basis of these effects can be understood because a Pro residue, given its rigidity and constrained backbone dihedral angles as noted before (*51*), is not compatible with the sharp turn that the A-loop makes at this position in the I_2 state, whereas it can be tolerated in the open conformation of the A-loop observed in Abl I_1 (Fig. 3D).

Several imatinib-resistance mutations cluster in the P-loop, including G269E, Q271H, Y272H, and E274V (G, Gly; E, Glu; Q, Gln; V, Val) (Fig. 3B). It has been suggested that CML patients with P-loop mutations have a poorer prognosis (*60*). Tyr^{272} is a key residue in stabilizing the stretched P-loop conformation in the Abl I_2 state by packing against Phe^{302} of the αC helix and using its hydroxyl group to form a hydrogen bond with the side chain of Glu^{305} (Fig. 2C). The transition of the P-loop from the kinked conformation observed in the active state to the stretched conformation in the I_2 state is necessary to accommodate the closed conformation of the A-loop in I_2. Quantitative NMR CEST and ^1H-^{13}C–correlated experiments show that the Y272H substitution decreases the population of the I_2 state ~7-fold (Fig. 3E), accounting for the ~10-fold decrease in the affinity of imatinib for Abl^{Y272H} (Fig. 3C and fig. S9). His^{272} is not well poised to form a hydrogen bond with Glu^{305}, and its aromatic π stacking with Phe^{302} is suboptimal compared with Tyr^{272}, resulting in destabilization of the stretched P-loop conformation and decrease of the Abl I_2 population.

Glu^{274} forms an ion pair with Lys^{293}, thereby stabilizing the stretched conformation of the P-loop in the I_2 state, and is frequently mutated to a Val in patients that have developed imatinib resistance (Fig. 3F). Because the singly mutated Abl^{E274V} variant is not stable enough for CEST studies, we measured the effect of the E274V substitution using the $Abl^{G269E/M309L/T408Y}$ (referred to as Abl^{I2M}) variant, which populates primarily the I_2 state (p_{I2} ~90%), by ^1H-^{13}C–correlated spectra. The data showed that the loss of the Glu^{274}-Lys^{293} ion pair results in destabilization of the Abl I_2 state by ~1.4 kcal mol^{-1}, which translates to a several-fold decrease in its population, thus explaining the lower affinity of imatinib for Abl^{E274V} (Fig. 3, C and F).

The F378V (F, Phe) imatinib-resistant mutation (Fig. 3B) decreases binding to the inhibitor ~4-fold (Fig. 3C and fig. S9). Consistent with this, NMR analysis of the Abl^{F378V} variant reveals that the F378V substitution decreases the population of the I_2 state ~4-fold (Fig. 3G). On the basis of NMR analysis, we surmise that Val^{378} exerts its effect by stabilizing the active state, thereby shifting the conformational ensemble toward the active state and depleting the I_2 state (fig. S10A). Our NMR data also explain the imatinib-resistance phenotype of two additional mutations in the P-loop, G269E and Q271H, which exhibit a more complex behavior (fig. S10, B and C). Taken together, our findings illustrate how amino acid substitutions in different structural elements of the Abl kinase can use various mechanisms to destabilize the conformational state to which imatinib selectively binds, giving rise to drug resistance.

The role of the regulatory spine in activity regulation

A highly conserved structural feature in protein kinases is the so-called regulatory (also referred to as the hydrophobic) spine (*61*). Mutations in this spine can directly activate or suppress the catalytic activity of kinases (*62*–*64*); however, direct experimental evidence on the underlying mechanisms is lacking. In

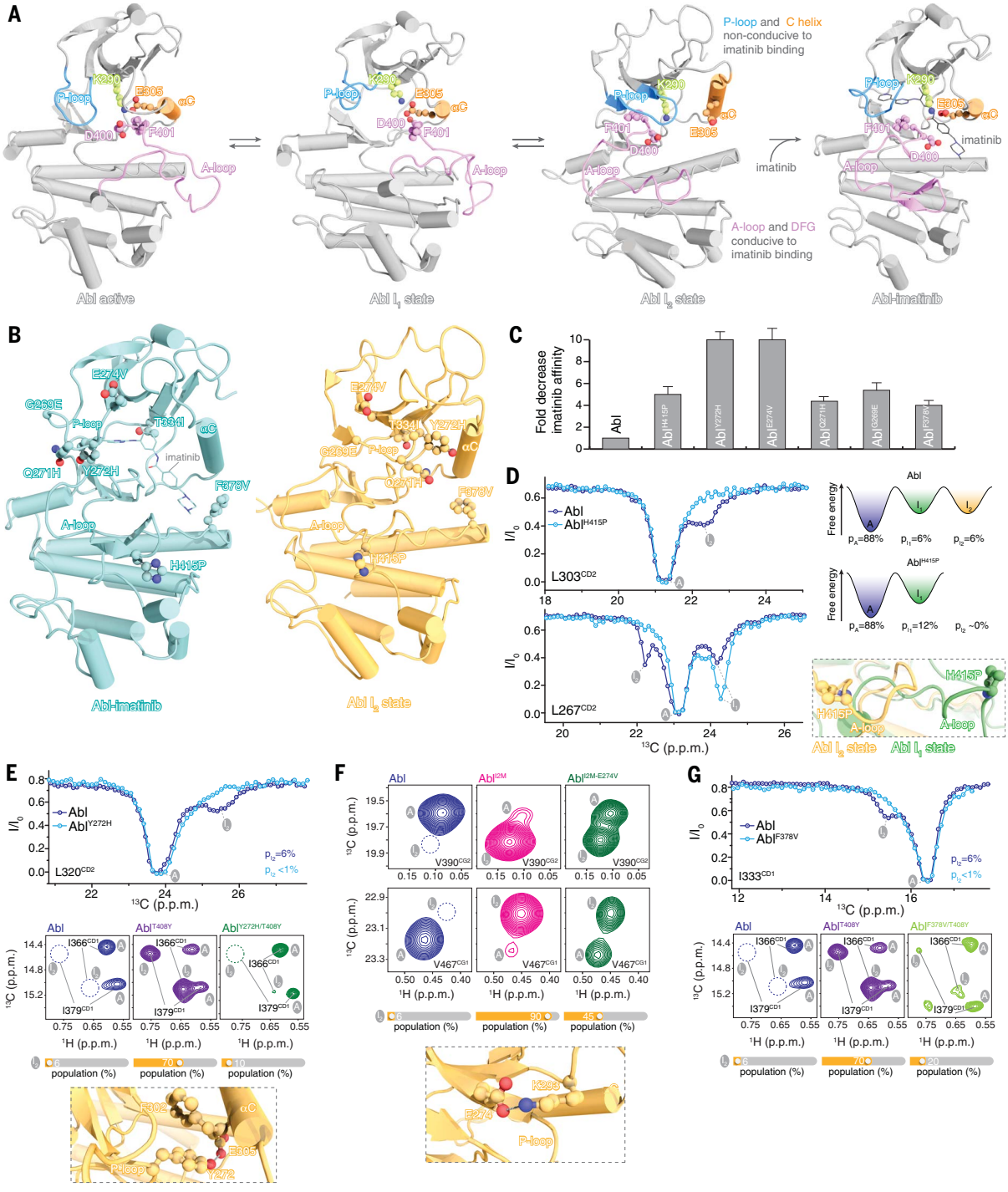


Fig. 3. Structural and energetic dissection of imatinib-resistance Abl variants. (**A**) Structures of the three Abl conformational states and binding of imatinib to the I_2 state highlighting the disposition of key structural elements. (**B**) Patient-derived imatinib-resistance mutants studied here shown on the structure of Abl in complex with imatinib and on the structure of the Abl I_2 state. (**C**) Change in affinity for imatinib by the Abl variants relative to wild-type Abl, as determined by isothermal titration calorimetry (fig. S9). Error bars are standard deviation determined from a triplicate. (**D**) Effect of the H415P substitution on the population of the Abl states measured by ^{13}C CEST experiments. The changes in the populations are indicated in a free-energy diagram. A magnified view of the mutation site suggests how the Pro substitution may destabilize the Abl I_2 state. (**E**) Effect of the Y272H substitution on the population of the Abl states measured by ^{13}C CEST using wild-type Abl as reference (top) and by ^{1}H-^{13}C–correlated experiments using the Abl^{T408Y} variant, which populates the Abl I_2 state at 70%, as reference (bottom). A magnified view of the mutation site indicates that substitution of Tyr^{272} will disrupt the hydrogen bond to Glu^{305}, thereby destabilizing the Abl I_2 state. (**F**) Effect of the E274V substitution on the population of the Abl states measured by ^{1}H-^{13}C–correlated spectra using the Abl^{I2M} variant as reference. A magnified view of the mutation site indicates that substitution of Glu^{274} will disrupt the hydrogen bond to Lys^{293}, thereby destabilizing the Abl I_2 state. (**G**) Effect of the F378V substitution on the population of the Abl states measured by ^{13}C CEST using wild-type Abl as reference (top) and by ^{1}H-^{13}C–correlated spectra using the Abl^{T408Y} variant as reference (bottom). For additional data and discussion, see fig. S10.

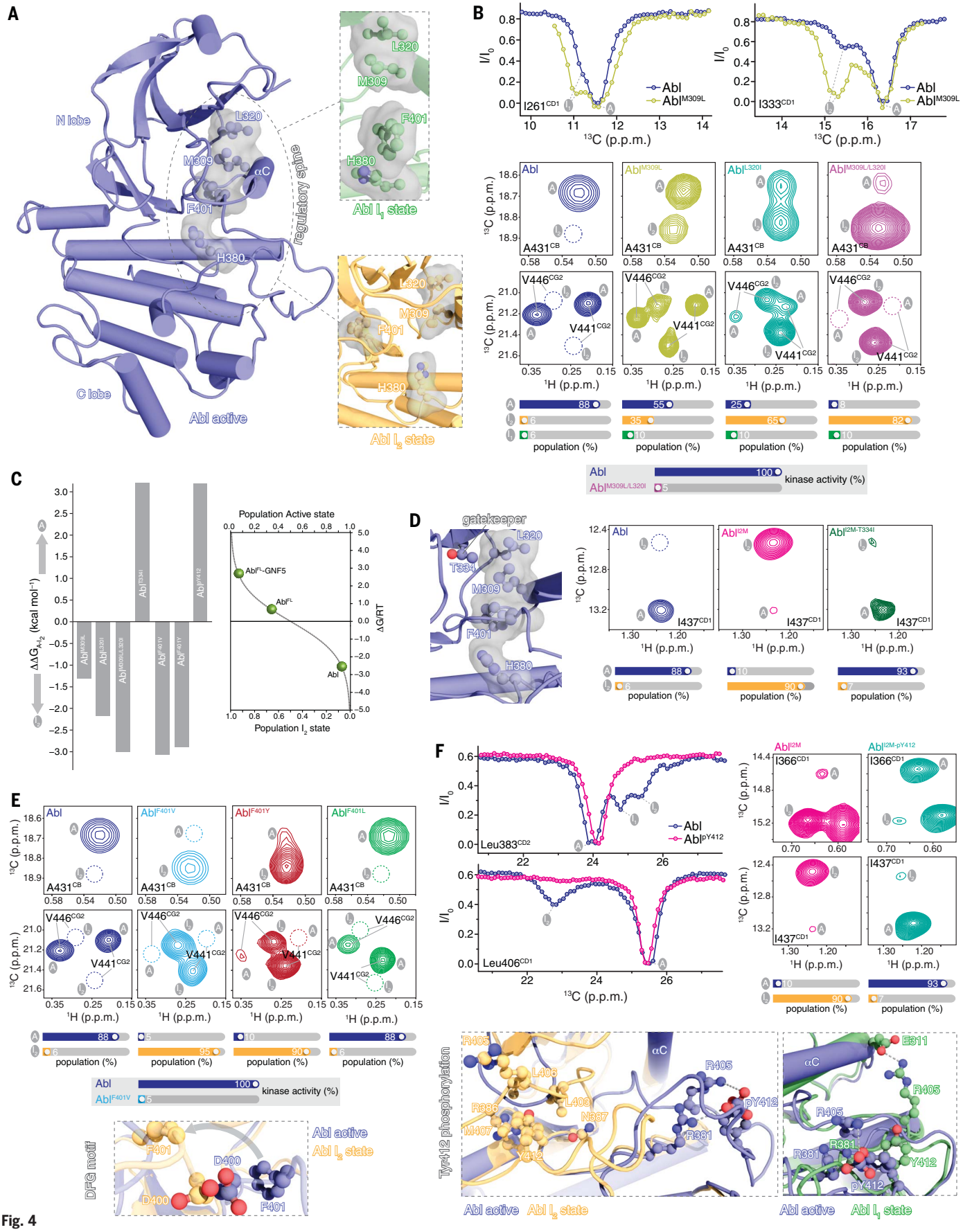

A
N lobe
C lobe
Abl active
αC
regulatory spine
L320
M309
F401
H380
Abl I1 state
Abl I2 state
B
I/I0
Abl
AblM309L
I261CD1
I333CD1
13C (p.p.m.)
AblL320I
AblM309L/L320I
A431CB
V446CG2
V441CG2
1H (p.p.m.)
population (%)
kinase activity (%)
C
ΔΔGA-I2 (kcal mol−1)
AblT334I
AblF401V
AblF401Y
AblF412
Population Active state
AblFL-GNF5
AblFL
Abl
ΔG/RT
Population I2 state
D
gatekeeper
T334
I437CD1
AblI2M
AblI2M-T334I
E
AblF401V
AblF401Y
AblF401L
DFG motif
D400
F401
Abl active
Abl I2 state
F
AblpY412
Leu383CD2
Leu406CD1
AblI2M-pY412
I366CD1
R405
L406
R386
M407
N387
R381
Y412
pY412
αC
E311
Tyr412 phosphorylation

Fig. 4

Fig. 4. Quantitative dissection of the effect of the regulatory spine, the gatekeeper, the DFG motif, and phosphorylation on the Abl conformational ensemble. (**A**) Structure of the active state of Abl highlighting the regulatory spine, made up by Met^{309}, Leu^{320}, His^{380}, and Phe^{401}. The disposition of the regulatory spine is also shown for Abl states I_1 and I_2. (**B**) Effect of amino acid substitutions at the regulatory spine on the populations of the Abl conformational states, as measured by NMR. Because the k_{ex} rates of the transitions (Fig. 1D) are slow on the NMR chemical shift time scale, the resonances of the individual conformational states can be seen in ^{1}H-^{13}C–correlated spectra if their population is above ~5%. Thus, in addition to ^{13}C CEST, ^{1}H-^{13}C–correlated experiments were also used to quantitate the populations, especially for Abl variants that are not stable over the period of time required for CEST experiments (Materials and methods). There is excellent agreement in the population measurements by CEST and ^{1}H-^{13}C–correlated experiments. The labeling scheme of the Abl variants is as shown in fig. S1. Determination of the kinase activity (fig. S11) shows that the $Abl^{M309L/L320I}$ variant inhibits Abl. (**C**) Energy contribution to the active and I_2 states of Abl by the indicated variants. Negative values of free-energy change, $\Delta\Delta G$, indicate increased stability of the Abl I_2 state, whereas positive values indicate increased stability of the Abl active state. A graph is included where the populations of the active and I_2 states are plotted as a function of the associated free energy, $\Delta G/RT$, where R is the gas constant and T is the temperature. A 0.6 kcal mol^{-1} change in ΔG corresponds to a change by 1 unit in $\Delta G/RT$ at room temperature (*77*). The populations of the two states for Abl (isolated kinase domain) and full-kinase Abl^{FK} are shown to help assess the effect that each one of the variants has on the activation or inhibition of Abl. (**D**) Effect of the Thr^{334}-to-Ile substitution at the gatekeeper position on the populations of the Abl states. Given that Abl exists predominantly in the active state (p_A ~ 88%), to measure the full effect, we used the Abl^{I2M} variant ($Abl^{G269E/M309L/T408Y}$), which populates predominantly the I_2 state (p_{I2} ~ 90%). (**E**) Effect of amino acid substitutions of the Phe residue in the DFG motif on the populations of the Abl states. Val or Tyr substitution results in an Abl kinase that adopts primarily the I_2 state. The low kinase activity of Abl^{F401V} is consistent with the NMR findings. (**F**) Effect of Tyr^{412} phosphorylation in the A-loop on the populations of the Abl states. Tyr^{412} phosphorylation eliminates both the I_1 and I_2 states and stabilizes the active state, as evidenced by ^{13}C CEST experiments. To measure the full effect, we used the Abl^{I2M} variant as discussed in (D). The magnified views of the superposition of the structure of each of the inactive states on the structure of the active state provides mechanistic insight into the effect.

Abl, the four residues making up the regulatory spine are Met^{309}, Leu^{320}, His^{380}, and Phe^{401} (Fig. 4A). We used NMR to obtain structural and thermodynamic insight into how mutations in the regulatory spine may remodel the ensemble of the active and the inactive states in Abl. Our data show that the spine is fully assembled in the active state but broken in both the I_1 and I_2 states owing to the DFG motif being in the "out" conformation (Fig. 4A). The structural deviation is more pronounced in the I_2 state than in the I_1 state because of the move of Phe^{401} toward the hydrophobic pocket in the nucleotide binding site. ^{13}C CEST and ^{1}H-^{13}C–correlated NMR experiments showed that substitution of Leu, a residue more frequently seen in kinases (*65*), for Met at position 309 increases the population of I_1 and I_2 from 6% each to 10 and ~35%, respectively, and hence increases the total population of inactive states ~4-fold (Fig. 4B). Substitution of the bulkier Ile for Leu at position 320 has an even greater effect, with the I_2 state population increasing to 65% (Fig. 4B). The double M309L/L320I substitution switches almost the entire ensemble toward the I_2 state (p_{I2} ~ 82%) and decreases the active state population to 8%, essentially deactivating the kinase, as confirmed by kinase assays (Fig. 4B and fig. S11). The energetic contributions of the M309L and L320I substitutions toward stabilizing the inactive states are 1.3 and 2.2 kcal mol^{-1}, respectively (Fig. 4C).

Although residues with higher hydrophobicity would be expected to promote the assembly of the spine, and thus stabilize the active state, our data show that this is not the case. Both the Met-to-Leu (position 309) and the Leu-to-Ile (position 320) substitutions are toward more hydrophobic residues (*66*), and both stabilize the inactive state. It is likely that in addition to hydrophobicity, the size and shape of the residue side chain at these positions are important in determining the packing and thus the overall assembly of the spine. It is of interest that a Met at the residue corresponding to the Abl 309 position is conserved in ~21% of protein kinases, whereas a Leu is conserved in ~54% of them. At the residue corresponding to the Abl 320 position, a Leu is conserved in ~54% and an Ile in ~8% of them (*65*). Thus, even conservative mutations that may cause only subtle changes in the hydrophobicity and size of the amino acids in the regulatory spine can give rise to pronounced changes in the population of the active and inactive states, thereby strongly regulating kinase activity. This finding may explain why the regulatory spine is a hotspot of oncogenic mutations in various kinases (*63*).

Effect of the gatekeeper residue

The residue at the gatekeeper position, which controls accessibility to the hydrophobic region of the ATP pocket in protein kinases, is important for at least two reasons: (i) Mutations to bulky residues, such as Ile and Met, at the gatekeeper position (*67*, *68*) stimulate the intrinsic kinase activity and give rise to drastically increased cell transforming and oncogenic activity (*62*, *69*); and (ii) mutations to bulky residues confer drug resistance in some kinases (*24*, *70*). The mechanistic basis for the gatekeeper-induced kinase activation is not known. In Abl, the strongest effect reported is for the gatekeeper Thr^{334} residue substituted by Ile (*62*, *71*). We characterized Abl^{T334I} by NMR, and the analysis showed that the T334I substitution shifts the equilibrium toward the active state. Because the Abl kinase domain already exists predominantly in the active state (p_A ~ 88%), it is experimentally challenging to quantitate this effect. To do this, we used the Abl^{I2M} variant, which populates primarily the I_2 state (p_{I2} ~ 90%) while the active state is only marginally stable (p_A ~ 10%) (Fig. 4D). T334I reverses the populations by strongly promoting the active state (p_A ~ 93%) (Fig. 4D). The data show that the Ile in the gatekeeper position stabilizes the active state by ~3.2 kcal mol^{-1} (Fig. 4C) and reveal that the basis for kinase activation is a shift of the conformational ensemble toward the catalytically active state. Presumably, the Ile exerts its effect by stabilizing the hydrophobic spine (*65*) through favorable contacts with their nonpolar side chains (Fig. 4D).

The role of the DFG motif in activity regulation

Mutations in the DFG motif have been identified as "drivers" in human cancers in several kinases (*72*). Although the highly conserved nature of the Asp residue in the DFG motif is explained by its key role in the chemical steps of the catalysis (*73*), the reason for the conservation of the Phe residue is not well understood (*74*). We thus prepared the Abl^{F401V} variant, which, as NMR data show, adopts almost exclusively (higher than 95%) the I_2 state (Fig. 4E). The F401V substitution in Abl flips the DFG motif from its preferable "in" state to the "out" state, which is stabilized by ~3.0 kcal mol^{-1} (Fig. 4C). In agreement, kinase assays showed that the catalytic activity of Abl^{F401V} is at least 20-fold lower than that of the wild-type protein (Fig. 4E and fig. S11). We observed a similarly strong shift toward the DFG-out state in the Abl^{F401Y} variant (p_{I2} ~ 90%; Fig. 4E). Thus, even a conservative substitution in this position can affect the ensemble of active and inactive conformational states, giving rise to impaired function. This observation is consistent with and may explain findings on other kinases. For example, p38α loses its activity on substitution of the Phe residue of the DFG motif to Tyr (*75*), and LRRK2, which features a Tyr in this position, is retained in an inhibited state and substitution by Phe activates the kinase (*76*).

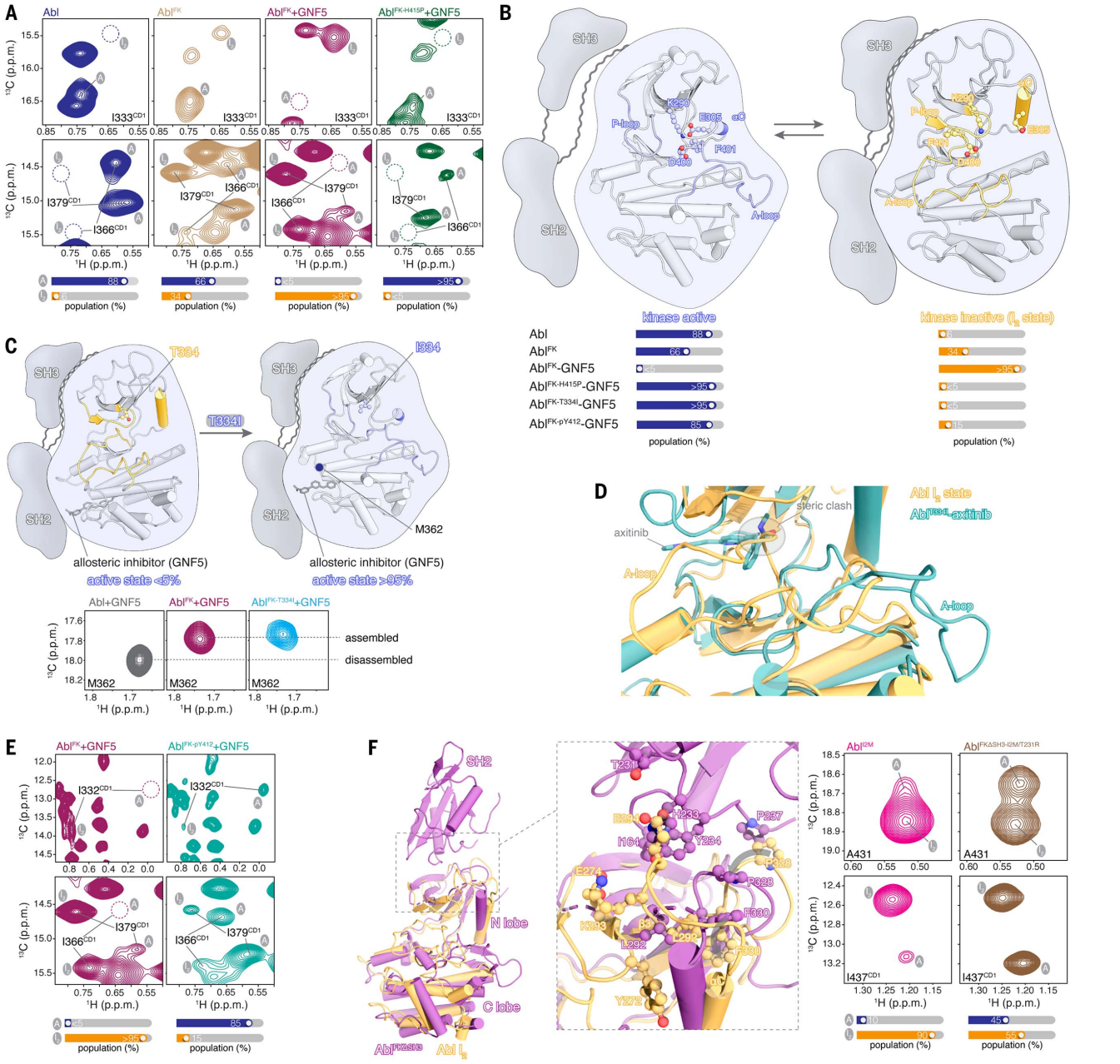


Fig. 5. Quantitative dissection of the effect of variants on the conformational ensemble of the full-length Abl kinase. (**A**) Effect of the SH3-SH2 regulatory module, the addition of the allosteric inhibitor GNF5, and the H415P substitution, as measured by NMR. Abl^{FK} corresponds to the SH3-SH2-KD Abl fragment. Because Abl^{FK} is not sufficiently stable for ^{13}C CEST experiments, we used ^{1}H-^{13}C–correlated spectra to quantitate the populations. For the isolated Abl kinase domain, we know from CEST experiments that the population of the I_1 state is 6%. However, although we can conclude from the ^{1}H-^{13}C–correlated spectra that the population of the I_1 state does not increase in Abl^{FK} and its variants, we cannot conclude if it is depleted. Thus, the populations reported for Abl^{FK} variants are only for the active and the I_2 states. (**B**) Schematic of Abl^{FK} showing the equilibrium between a disassembled conformation, wherein the regulatory module is not bound to the kinase domain, and an assembled conformation, wherein the regulatory module is docked onto the back of the kinase domain. In the disassembled conformation, the kinase adopts the active state, whereas in the assembled one, it adopts the I_2 state. The populations of the active and I_2 states in the Abl variants studied are indicated. (**C**) Schematic of Abl^{FK} summarizing the effect of the gatekeeper T334I substitution, which forces Abl to adopt the active state even when the regulatory module is docked onto the kinase. ^{1}H-^{13}C–correlated spectra showing the M362 methyl resonance. M362 is located at the interface between the SH2 and the kinase domain and provides a sensitive probe for the assembled conformation. (**D**) Superposition of the structures of the Abl I_2 state and the Abl^{T334I}-axitinib complex (PDB ID 4TWP) highlights the steric clash between the bound inhibitor and the A-loop in the I_2 state, which explains the higher affinity of the inhibitor for the T334I variant. (**E**) Effect of Tyr^{412} phosphorylation on the populations of the active and I_2 states of Abl^{FK} in complex with the allosteric inhibitor GNF5, as measured by NMR. (**F**) Effect of the extended conformation, wherein the SH2 domain docks on the top of the N-lobe, on the populations of the active and I_2 states of Abl, as measured by NMR. The structure of the Abl$^{FK\Delta SH3}$ fragment (purple; PDB ID 4XEY) is superimposed on the structure of the Abl I_2 state (orange).

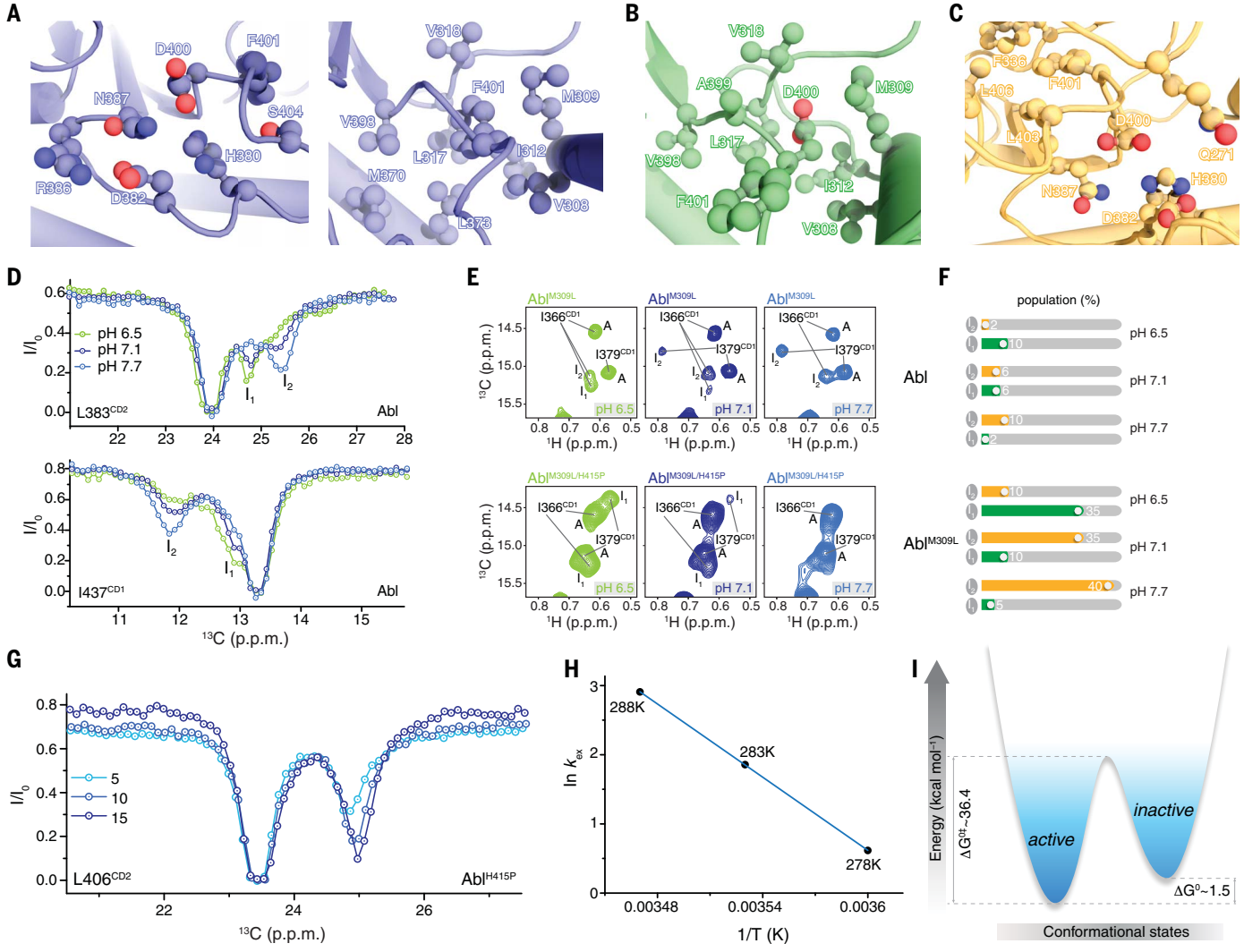


Fig. 6. pH effect on the Abl conformational ensemble and activation energy of the DFG transition. (**A** to **C**) Close-up view of the DFG motif conformation in the (A) active state, (B) I_1 state, and (C) I_2 state. The positioning of Asp^{400} of the DFG motif differs between the various conformational states. In the active state, Asp^{400} is exposed to the solvent, whereas Phe^{401} is buried inside a hydrophobic pocket. The DFG motif flips 180° as it transitions from the active to the I_1 state, with the two residues swapping positions: In the I_1 state, Asp^{400} is now positioned inside the hydrophobic pocket, whereas Phe^{401} is exposed to the solvent. Although the DFG motif remains in the "out" conformation in the I_2 state, the A-loop undergoes a major rearrangement, and as a result, Asp^{400} is found in a polar environment, similarly to the active state, and Phe^{401} is buried inside the hydrophobic pocket located in the nucleotide binding site. (**D**) ^{13}C CEST profiles of representative residues as a function of pH. The results show that lower pH (pH 6.5) selectively stabilizes the I_1 state, whereas higher pH (pH 7.7) selectively stabilizes the I_2 state. (**E**) ^{1}H-^{13}C–correlated spectra of Abl^{M309L} and $Abl^{M309L/H415P}$ variants as a function of pH. In $Abl^{M309L/H415P}$ at pH 7.7, there is no detectable inactive state because the H415P substitution eliminates the I_2 state and the higher pH depletes the I_1 state. (**F**) Populations of the I_1 and I_2 states as a function of pH. The relative populations of the I_1 and I_2 states fluctuate antagonistically, but the ratio of populations between the active state and the inactive states collectively is not affected by changes in pH within this range. (**G**) ^{13}C CEST profiles of Abl^{H415P} as a function of temperature. (**H**) Arrhenius plot of the k_{ex} measured by the CEST experiments in (G) for the three temperatures indicated, used to determine the activation energy of the DFG transition from the active to the I_1 inactive conformation. (**I**) Energy diagram indicating the activation enthalpy $\Delta G^{0\ddagger}$ (~36.4 kcal mol^{-1}) of the DFG flip from DFG-in to DFG-out and the free energy difference, ΔG^0, between the two states.

Although Abl I_2 is stabilized in the Abl^{F401V} and Abl^{F401Y} variants, the I_1 state is not. The energetic balance of the contacts of the residue at position 401 with the residues in the two hydrophobic pockets where Phe^{401} was observed to reside into in the active and the I_2 states (Fig. 2, A and C) will determine the relative populations of the two states and thus the activity of the kinase.

The Phe residue in the DFG motif is conserved in ~90% of kinases, but ~7% of them feature a Leu residue (*65*). In cAMP-dependent protein kinase, the Phe-to-Leu mutation has no effect on its catalytic activity (*65*). The NMR data of Abl^{F401L} show that this variant populates predominantly the active state, similarly to wild-type Abl (Fig. 4E), thus explaining why a Leu residue in this position can be tolerated.

The role of A-loop phosphorylation in kinase activation

Phosphorylation of Tyr, Thr, or Ser residues in the A-loop is a common mechanism for stimulating the activity of protein kinases (*38*), including Abl (*42*). We quantitated the effect of Abl Tyr^{412} phosphorylation ($pTyr^{412}$) on the conformational ensemble of active and inactive states. NMR CEST data show that in Abl^{pY412},

both inactive states I_1 and I_2 are depleted and only the active state is detected (Fig. 4F). Our structural data illustrate the mechanistic basis for this effect. Both the active and I_1 states feature an open A-loop conformation with Tyr^{412} exposed to the solvent. In the active state, the phosphate group of $pTyr^{412}$ forms ion pairs with Arg^{381} and Arg^{405} (Fig. 4F). In the I_1 state, although Arg^{381} is well poised to form a salt bridge with $pTyr^{412}$, Arg^{405} is pointing outward and interacts with Glu^{311} (Figs. 2B and 4F). Phosphorylation of Tyr^{412} causes Arg^{405} to flip inward to form a salt bridge with $pTyr^{412}$, thereby destabilizing the I_1 state.

In the I_2 state, Tyr^{412} binds as a pseudosubstrate and blocks the substrate-binding site (Figs. 2C and 4F). The side chain of Tyr^{412} is positioned in a hydrophobic pocket that is lined by Leu^{403}, Leu^{406}, and Met^{407}. Thus, phosphorylation of Tyr^{412} destabilizes the I_2 state by removing the negatively charged $pTyr^{412}$ from the pocket (Fig. 4F). To quantitatively measure the effect of Tyr^{412} phosphorylation on the energetics of the ensemble, we used the Abl^{I2M} variant. The population of I_2 in Abl^{I2M} is 90%, whereas it is only 10% in $Abl^{I2M\text{-}pY412}$ (Fig. 4F). Thus, phosphorylation of Tyr^{412} stabilizes the active state by ~3.0 kcal mol^{-1} (Fig. 4C) and can activate even Abl variants that exist predominantly in the inactive state.

Allosteric regulation of Abl kinase by modulation of the conformational ensemble

Similarly to the Src family kinases (*3*), the docking of the SH3-SH2 regulatory module onto the kinase domain of Abl suppresses its activity (*52*). Because no structural data are available on an SH3-SH2-kinase Abl fragment without an inhibitor bound, the structural changes to the kinase domain elicited by the SH3-SH2 module are not known. We prepared and studied by NMR Abl fragments that encompass the SH3-SH2 module and the kinase domain, which, for simplicity, we refer to as the full kinase (Abl^{FK}). The presence of the SH3-SH2 module in Abl^{FK} increases the population of the I_2 state ~6-fold, from 6% in the isolated kinase domain to 34% in Abl^{FK} (Fig. 5A). As shown previously (*77*), only ~40% of the molecules in Abl^{FK} are found in the fully assembled state—that is, where the SH3-SH2 module is docked onto the back of the kinase—whereas ~60% of the molecules adopt a disassembled conformation. The allosteric inhibitor GNF5, which was previously shown to promote the assembled conformational state (*32*, *77*), increased the population of the assembled Abl^{FK} state to ~95% (*77*). NMR data on Abl^{FK}-GNF5 indicate that GNF5 binding stabilizes the I_2 state so that its population rises to 95% (Fig. 5A). Thus, our data clearly establish that allosteric stabilization of the I_2 state by the SH3-SH2 module is the structural basis for inhibition in Abl^{FK} (Fig. 5B). The I_1 inactive state is not stabilized by the docking of the SH3-SH2 module. The H415P substitution, which abrogates the I_2 state and promotes the I_1 state in the isolated kinase domain (Fig. 3D), shifts the equilibrium markedly toward the active state in $Abl^{FK\text{-}H415P}$, thereby effectively counteracting the autoinhibitory mechanism (Fig. 5, A and B). This explains the puzzling finding that the H415P substitution stimulates Abl kinase activity (*58*).

Quantitative information on the conformational ensemble of Abl^{FK} provides a tool to assess the effects that various factors—such as the gatekeeper, phosphorylation, and allosteric inhibitors—exert on the kinase. As shown above for the isolated Abl kinase domain, the T334I gatekeeper substitution stimulates Abl activity by shifting the equilibrium toward the active state (Fig. 4, C and D). In Abl^{FK}, the T334I substitution promotes the active state even in the presence of the GNF5 allosteric inhibitor (Fig. 5C). Therefore, this single mutation is capable of activating Abl even when the kinase is held almost entirely in its assembled state (Fig. 5C). This finding is consistent with and explains the observation that the T334I variant markedly stimulates kinase activity even in the myristoylated form of Abl (*62*, *69*). Our data also explain why allosteric inhibitors are not effective against the T334I gatekeeper substitution (*78*, *79*), even though Ile^{334} is located away from the allosteric pocket. Our finding that $Abl^{FK\text{-}T334I}$ adopts predominantly the active state is consistent with and accounts for the ~10-fold higher potency of axitinib, which binds selectively to the active state of Abl kinase, in inhibiting $Bcr\text{-}Abl^{T334I}$ compared with Bcr-Abl (*48*), because the latter exists predominantly in the I_2 state and is not compatible with axitinib binding (Fig. 5D). Similar to the effect exerted by T334I, phosphorylation of the A-loop at Tyr^{412} counteracts the strong autoinhibitory mechanism in $Abl^{FK\text{-}GNF5}$ by destabilizing the I_2 state (Fig. 4F) and promoting the active state (Fig. 5, B and E). This finding explains the much lower affinity of imatinib for the phosphorylated ($pTyr^{412}$) Abl (*80*).

An alternative to the fully assembled, downregulated state of Abl^{FK} is an experimentally observed conformational state wherein the SH2 domain docks on the top of the N-lobe, thus giving rise to an extended conformation (*77*, *81*–*83*). This structural arrangement is known to stimulate the activity of Abl and has been associated with increased leukemogenic activity (*58*, *82*, *84*). The structural basis for this activity stimulation is not understood. To quantitatively assess the effect, we studied by NMR the $Abl^{FK\Delta SH3\text{-}I2M/T231R}$ variant; this variant includes the I2M triple mutant that populates predominantly the I_2 state; lacks the SH3 domain, which promotes the assembled conformation; and features the T231R (R, Arg) substitution, which is thought to stabilize the interface between the SH2 and the N-lobe (*58*, *77*). The results show that docking of SH2 on the N-lobe of the kinase promotes the active state by a factor of ~4, corresponding to an energetic contribution of ~1.2 kcal mol^{-1} (Fig. 5F). In the context of Abl^{FK}, stabilization of the SH2–N-lobe interface is thus expected to increase the population of the active state from ~35 to ~80% (Fig. 4C). Superposition of the structure of an Abl fragment crystallized in the extended conformation (*83*) with the structure of the Abl I_2 state shows that the surface of the N-lobe in the I_2 state is not compatible with SH2 binding (Fig. 5F). A steric clash between the N-lobe Glu^{294} and the SH2 His^{233} and Tyr^{234} forces the loop connecting the αC helix and β3 strand to retract, thereby disrupting the ion pair between Glu^{274} and Lys^{293} (Fig. 5F), which is essential for the stabilization of the stretched conformation of the P-loop observed in the Abl I_2 state (Fig. 3F). Taken together, our data show that binding of the SH2 domain on the top of the N-lobe disrupts the αC helix–P-loop conformation in the I_2 state and thus shifts the equilibrium toward the fully active state. Of note, the T231R mutation has been identified as a patient-derived imatinib-resistant variant. Our data explain how the T231R substitution, which strengthens the SH2–N-lobe interface, confers resistance to imatinib (*58*).

pH changes redistribute the populations between the inactive states

Given the distinct environment of Asp^{400} in the three conformational states (Fig. 6, A to C), we asked whether subtle changes in the electrostatic environment coupled to variations in Asp^{400} pK_a [negative log of the acid dissociation constant (K_a)] could affect the conformational ensemble and thus the kinase activity. We measured the populations of the three conformational states of Abl as a function of pH (Fig. 6, D and E) and found that even relatively small changes in the pH can have a substantial effect on the relative stability of the two inactive states. Specifically, although at pH 7.1 the I_1 and I_2 states have the same population, at pH 6.5, the population of the I_1 state is several-fold higher than that of the I_2 state (Fig. 6F). Conversely, at pH 7.7, the I_2 state is much more stable than the I_1 state. Because of the high energy (*85*) required to bury the charged side chain of Asp^{400} in the hydrophobic pocket in the I_1 state, Asp^{400} needs to be protonated (*86*). This is in agreement with our observation that lower pH values favor the I_1 state. Of interest, pH changes within this range (6.5 to 7.7) do not seem to affect the relative stability of the active versus the inactive states

but rather only redistribute the population within the two inactive states.

Activation energy of the DFG transition

The kinetics of the DFG flip determine how fast a kinase is activated or inhibited (*86*, *87*) as it toggles between active and inactive states in response to stimuli. The kinetics of the DFG flip may also determine the rate of adenosine diphosphate (ADP) release (*74*), which is typically the rate-limiting step in kinase catalysis (*73*). To measure the activation energy of the DFG transition, which governs its kinetics, we measured the exchange rate (k_{ex}) between the active state and the I_1 state as a function of temperature (Fig. 6, G and H). The data showed that the activation energy of the DFG transition from the active to the inactive state is ~36 kcal mol^{-1} (Fig. 6F). This large energy barrier indicates that the DFG flip in Abl kinase entails pronounced structural rearrangement in the transition state and explains why the DFG flip is associated with intrinsically slow kinetics.

Discussion

Here, we describe in structural and energetic terms the transition of Abl kinase between its active and two inactive conformational states. All of these states are intrinsic to the isolated kinase domain, and they determine how the kinase responds to regulatory stimuli, substrate and ligand binding, allosteric interactions, and posttranslational modifications. Abl kinase populates two distinct inactive states that are quite different from each other. Whereas the DFG motif is in the "out" conformation in both the I_1 and I_2 states, the A-loop and the αC helix adopt very different conformations. Thus, there are various mechanisms and pathways for a kinase to switch to an inactive conformation.

The Abl I_2 state, which adopts a fully inactive conformation, is selectively stabilized in the full-length kinase (Fig. 5B). The biological importance of the I_1 state is not apparent. Although our data suggest that I_1 lies in the pathway between the active and the I_2 state, depletion of the I_1 state has no effect on the population of the I_2 state (Fig. 2E) and therefore is likely not an obligatory intermediate species. Moreover, a single mutation (H415P) that abrogates the I_2 state converts the kinase to its fully active conformation, with the I_1 state being very poorly populated even in the full length Abl kinase. That is, the inhibitory mechanism in Abl imposed by the docking of the regulatory module onto the kinase operates only on the I_2 state and not on the I_1 state (Fig. 5, A and B). Similarly, mutations in the DFG motif that inhibit the kinase exert their effect by selectively stabilizing the I_2 state and not the I_1 (Fig. 4E). It is possible that the I_1 state is a by-product of the evolutionary process that established the active and the I_2 state, the two functional states. Although conformational states such as the I_1 may not be associated with a known biological function, they could be leveraged for the design of selective inhibitors. Characterization of a number of kinases that use the approaches described here could provide evidence of the existence of low-populated inactive conformational states in the kinome.

The discovery and structural insight into the Abl I_2 state as the one to which imatinib selectively binds provides the mechanistic basis for explaining a number of patient-derived imatinib-resistance variants located at sites remote to imatinib. Our data demonstrate that all of these amino acid substitutions confer resistance to imatinib by depleting the population of the I_2 state. The I_2 state can be destabilized by altering the arrangement of any of several structural elements in this state—such as the A-loop, the P-loop, and the αC helix—either directly or allosterically (Figs. 4 and 5F and fig. S10). These mutations decrease the affinity not only of imatinib but also of other inhibitors—such as nilotinib, ponatinib, and rebastinib (*28*)—that also select the Abl I_2 state (fig. S4). The structural data of the Abl I_2 state reported here provide an opportunity for optimizing imatinib analogs for higher affinity.

A question that remains largely unaddressed is how many different inactive conformations intrinsic to kinases are present in the kinome (*18*). This is important not only for better understanding the variety of mechanisms used by kinases to regulate their activity but also for the design of selective inhibitors. Only a limited number of unliganded kinases have been crystalized in an inactive conformation, and only in a fraction of them are the intact A-loop, P-loop, and αC helix all visible (*65*). Among kinases with a known structure, the Abl I_1 state is distinctive. The Abl I_2 state is similar to the structure that the insulin receptor kinase (IRK) adopts in its inactive, nonphosphorylated form (*88*). Superposition of the structures of the Abl I_2 state with that of IRK highlights the notable similarities in the disposition of the A-loop, the DFG motif, the P-loop, and the αC helix (fig. S12A). Structural and sequence analyses of the two kinases indicated that the residue at position 408 in Abl (1158 in IRK) may be crucial in stabilizing the closed conformation of the A-loop. The single T408Y substitution in Abl increases the I_2 state population by more than an order of magnitude, from 6 to 70% (fig. S6). IRK has a Tyr in this position (fig. S12A). Four more receptor kinases—TrkA, TrkB, IGF1R, and MuSK—adopt a similar structure in their inactive state (fig. S12B). All of them feature a Tyr residue in this position. Conversely, three receptor kinases—FLT3, KIT, and FMS—that also adopt a similar structure in their inactive state do not bear a Tyr in this position; however, all of these three kinases require the presence of the juxtamembrane (JM) domain to stabilize their inactive state (fig. S12B). It should be noted that these receptor tyrosine kinases must adopt an inactive conformation in their resting state (*89*) and thus have evolved to stabilize such a conformation. Taken together, these findings underscore two key observations. First, a structurally similar inactive state among different kinases can be elicited by various mechanisms. For example, Abl accomplishes this by the allosteric action of the regulatory module. Some kinases, such as IRK, feature a crucial Tyr residue in the A-loop that directly stabilizes the inactive state. Other kinases, such as KIT, require the docking of the JM domain against the A-loop. Second, several kinases, both in the receptor and nonreceptor tyrosine kinase families, share a structurally similar inactive state. This observation suggests that there might be a limited number of structurally divergent inactive states intrinsic to kinases (*18*). More structural data on the intrinsic inactive conformations of kinases in the absence of inhibitors are needed to discover the full repertoire of their architecture and the mechanisms used to induce these conformations.

Our findings reveal additional mechanisms, other than allosteric interactions by intramolecular domains and direct binding of inhibitory proteins, that kinases use to induce their inactive states. The regulatory spine can switch the kinase between its active and inactive states even with conservative changes in the amino acid residues that make up the spine. In Abl, the double substitution M309L/L320I in the spine switches the kinase from a fully active state (p_A ~ 90%) to an almost fully inactive state (p_A ~ 8%) (Fig. 4B). This finding explains why many oncogenic mutations map on the regulatory spine (*63*). Another mechanism for a kinase to stabilize its inactive state is to feature an amino acid other than Phe in the DFG motif (Fig. 4E). Finally, our data highlight and explain how a single mutation at the gatekeeper position can completely neutralize autoinhibitory mechanisms to give rise to a fully activated form of a kinase.

Materials and methods

Expression and preparation of proteins

Abl constructs used in this work were cloned from full-length Abl isoform 1b. The coding sequences were cloned into the pET16b vector for expression as maltose binding protein (MBP)–His_6 fusion proteins including a tobacco etch virus (TEV) protease cleavage site. Various Abl mutants were generated using the QuikChange site-directed mutagenesis kit (Agilent), and their sequences were confirmed by DNA sequencing. The following Abl constructs were prepared: Abl^{FK} (residues 1 to 534 or 83 to 534), Abl kinase domain (referred to as Abl; residues 248 to 534), Abl^{G269E}, Abl^{Q271H},

Abl^{Y272H}, Abl^{M309L}, Abl^{E311K}, Abl^{L320I}, Abl^{T334I}, Abl^{F378V}, Abl^{F401V}, Abl^{F401L}, Abl^{F401Y}, Abl^{T408Y}, Abl^{H415P}, $Abl^{G269E/R405E}$, $Abl^{G269E/M309L}$, $Abl^{Y272H/T408Y}$, $Abl^{M309L/E311K}$, $Abl^{M309L/L320I}$, $Abl^{M309L/F401L}$, $Abl^{M309L/T408Y}$, $Abl^{M309L/H415P}$, $Abl^{E311K/T408Y}$, $Abl^{L320I/T408Y}$, $Abl^{F378V/T408Y}$, $Abl^{L389M/T408Y}$, $Abl^{T408Y/H415P}$, $Abl^{G269E/M309L/T334I}$, $Abl^{M309L/L320I/T334I}$, $Abl^{G269E/M309L/T408Y}$, $Abl^{G269E/E274V/M309L/T408Y}$, $Abl^{FK\text{-}M309L}$, $Abl^{FK\text{-}T334I}$, $Abl^{FK\text{-}T408Y}$, $Abl^{FK\text{-}H415P}$, $Abl^{FK\Delta SH3}$ (residues 138 to 534), $Abl^{FK\Delta SH3\text{-}T231R}$, and $Abl^{FK\Delta SH3\text{-}T231R/G269E/M309L/T408Y}$. Abl constructs were expressed and purified as described previously (*77*). Unlabeled protein samples were produced in cells grown in Luria-Bertani (LB) medium at 37°C in the presence of ampicillin (100 μg ml^{-1}) to an absorbance at 600 nm (A_{600}) of 0.8. Protein induction was induced by the addition of 0.2 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), and cells were allowed to grow for 48 hours at 16°C. Cells were harvested at A_{600} ~ 1.8 and resuspended in lysis buffer [50 mM Tris-HCl, 500 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM β-mercaptoethanol (BME), pH 8]. Cells were disrupted by a high-pressure homogenizer and centrifuged at 50,000*g*. Proteins were purified using Ni Sepharose 6 Fast Flow resin (GE Healthcare), followed by tag removal by TEV protease at 4°C (incubation for 16 hours) and gel filtration using Superdex 75 16/60 or 200 16/60 columns (GE Healthcare). Protein concentration was determined spectrophotometrically at 280 nm using the corresponding extinction coefficient. CrkII was cloned into the pET42a vector, and expression and purification were performed as previously described (*90*). Hck kinase used for Abl Tyr^{412} phosphorylation was cloned into the pET16b vector and was expressed and purified using a similar protocol to Abl.

Inhibitors

All Abl inhibitors (Selleck Chemicals) used in this study were solubilized in DMSO-D_6 (Cambridge Isotope Laboratories) at a concentration of 10 to 50 mM and titrated to Abl solutions from these concentrated stocks.

Isothermal titration calorimetry experiments

Isothermal titration calorimetry (ITC) experiments were performed on MicroCal iTC200 or Auto-iTC200 calorimeters (Malvern Instruments Inc.). ITC titrations for each Abl-inhibitor pair were typically performed at 25°C in 25 mM sodium phosphate buffer (pH 7.1) including 75 mM NaCl and 2.0 mM tris(2-carboxyethyl) phosphine (TCEP). Proteins were purified by size exclusion chromatography using corresponding ITC buffers before use. Concentrations of protein and inhibitors were measured spectrophotometrically and by weight, respectively. Proteins were placed in the cell while the inhibitors were in the syringe, with concentrations in the ranges of 8 to 60 μM and 80 to 650 μM, respectively. The small quantity of dimethyl sulfoxide (DMSO) remaining in the inhibitor buffer after the transfer from the stock solution was matched in the protein buffer. All solutions were filtered using membrane filters (pore size, 0.45 μm) and thoroughly degassed. The data were fitted with Origin 7.0 (OriginLab Corporation).

NMR sample preparation

Isotopically labeled samples for NMR studies were prepared by growing the cells in minimal (M9) medium. Cells were typically harvested at A_{600} ~ 1.2. U-[^{13}C,^{15}N]-labeled samples were prepared for the backbone assignment by supplementing the growing medium with $^{15}NH_4Cl$ (1 g $liter^{-1}$) and $^{13}C_6$-glucose (2 g $liter^{-1}$) (CIL and Isotec). The ^1H-^{13}C methyl-labeled samples were prepared as described (*91*, *92*). For Phe and Tyr selective labeling, 50 mg $liter^{-1}$ of U-[^2H] ε-^1H,^{13}C-Phe and Tyr were used. NMR samples of Abl variants were typically prepared in 25 mM sodium phosphate buffer (pH 7.1) containing 75 mM NaCl and 3.0 mM BME, in 100% ^2H_2O. The concentration of protein samples ranged from 30 to 300 μM. Abl-inhibitor complexes for NMR studies were prepared by mixing the protein at low concentration (~10 μM) with the inhibitor in the NMR buffer, followed by concentration through ultrafiltration.

NMR spectroscopy

All NMR experiments were acquired on Bruker 1.1-GHz and 900-, 850-, 800-, 700-, and 600-MHz spectrometers equipped with cryogenic probes. Spectra of unliganded Abl were typically collected at 10°C, whereas spectra for Abl-inhibitor complexes were collected at 10° or 20°C. All NMR data were processed using NMRPipe (*93*) and analyzed using NMRFAM-Sparky (*94*). Backbone ^1H, ^{15}N, and ^{13}C resonance assignment for Abl variants was achieved using 3D HNCACB, CBCA(CO)NH, HNCA, and HNCO experiments. Side-chain methyl and aromatic resonances were assigned using 3D-^{13}C,^{15}N-NOESY-HMQC, and ^{13}C-HMQC-NOESY-HMQC spectra (*95*) aided by the MAGIC software (*96*).

CEST experiments

$^{13}CHD_2$-CEST experiments (*40*, *97*) were recorded on Bruker 1.1-GHz and 900- and 850-MHz spectrometers at 10°C. Protein samples were prepared in 25 mM sodium phosphate (pH 7.1), 75 mM NaCl, 3.0 mM BME, and 100% D_2O. Protein concentration for CEST experiments ranged from 0.15 to 0.3 mM. The labeling scheme for the methyl-bearing residues was as follows: (i) [U-^2H; Leu,Val-$^{13}CHD_2$/$^{13}CHD_2$], (ii) [U-^2H; Ileδ1-$^{13}CHD_2$], (iii) [U-^2H; Met-$^{13}CHD_2$], (iv) [U-^2H; Ala-$^{13}CHD_2$], and (v) [U-^2H; Thr-$^{13}CH_3$]. A series of 2D spectra in a pseudo-3D mode were recorded with a mixing time (T_{mix}) of 500 ms and weak B_1 radio-frequency fields of 15 and 25 Hz applied at the ^{13}C dimension in steps of 20 to 25 Hz. A recycle time of 2.0 s was used in all experiments. Peak intensities were measured using NMRFAM-Sparky (*94*). CEST profiles were made by plotting the intensity ratios with and without the T_{mix} period.

CEST data fitting

CEST data were analyzed using ChemEx (https://github.com/gbouvignies/chemex), which numerically simulates the experiment by solving the Bloch-McConnell equations and minimizes the standard χ^2 equation to get best-fit exchange parameters (*41*, *97*). Errors in peak intensity for the ChemEx input were measured from the background noise of the spectra. Uncertainties in the exchange parameters were estimated by Monte Carlo simulations. Fitting data in a three-state exchange process is typically more challenging. For this reason, we sought to assign the individual minor dips to the corresponding excited state (E_1 and E_2). The strong and opposite dependence of the populations of the two excited states on pH facilitated this assignment, further confirmed by the CEST data on the Abl^{H415P} variant, which eliminates the I_2 state. We first proceeded to fit the CEST data for the Abl^{H415P} variant in which the exchange process is a two-state. Methyls were included in a global fitting to a two-state model of conformational exchange ($A \overset{k_{ex}}{\Leftrightarrow} E_1$) based on the following criteria: (i) there was no overlap with other peaks, (ii) a minor dip could be discerned in the CEST profile, and (iii) there was no effect by other local dynamics. The calculated k_{ex} and P_{E1} from fitting were 46.8 ± 4.3 s^{-1} and 12.1 ± 0.5%, respectively, with a reduced χ^2 (χ_{red}^2) value of 1.38. Next, we fitted the wild-type Abl CEST data by globally fitting to a three-state exchange model as follows:

$A \overset{k_{ex1}}{\Leftrightarrow} E_1 \overset{k_{ex2}}{\Leftrightarrow} E_2$ model 1

$E_1 \overset{k_{ex1}}{\Leftrightarrow} A \overset{k_{ex2}}{\Leftrightarrow} E_2$ model 2

$A \overset{k_{ex1}}{\Leftrightarrow} E_2 \overset{k_{ex2}}{\Leftrightarrow} E_1$ model 3

The χ_{red}^2 values obtained from models 1, 2, and 3 were 1.15, 1.29, and 1.41, respectively, suggesting that model 1 is the most appropriate model to fit the data. The exchange parameters obtained from model 1 were k_{ex1} = 12.3 ± 2.1 s^{-1}, k_{ex2} = 145.5 ± 22.6 s^{-1}, p_{E1} = 10.6 ± 1.2%, and p_{E2} = 9.6 ± 1.1%; the parameters obtained from model 2 were k_{ex1} = 14.7 ± 3.6 s^{-1}, k_{ex2} = 18.7 ± 2.7 s^{-1}, p_{E1} = 9.9 ± 1.7%, and p_{E2} = 7.0 ± 0.7%; and the parameters obtained from model 3 were k_{ex1} = 11.2 ± 2.4 s^{-1}, k_{ex2} = 136 ± 23.2 s^{-1}, p_{E1} = 21.2 ± 4.0%, and p_{E2} = 9.2 ± 1.2%. Populations of the excited states larger than ~10% were not consistent with the CEST profiles. Furthermore, we could not detect any peaks of the excited states in ^1H-^{13}C–correlated spectra of Abl recorded over a period of several

days, indicating lower populations. Because of the increased number of fitting parameters in a three-state exchange model, data fitting can yield multiple minima. We thus globally fit the data to both model 1 and model 2 by fixing k_{ex1} at 46.8 s^{-1}, as determined by the two-state CEST data fitting on Abl^{H415P}. The χ_{red}^2 values were 1.42 and 1.41 for model 1 and model 2, respectively. The exchange parameters obtained from model 1 were k_{ex1} = 46.8 ± 4.3 s^{-1}, k_{ex2} = 88.7 ± 13.5 s^{-1}, p_{E1} = 5.9 ± 0.1%, and p_{E2} = 6.1 ± 0.7%; and the parameters from model 2 were 46.8 ± 4.3 s^{-1}, k_{ex2} = 19.8 ± 2.9 s^{-1}, p_{E1} = 5.4 ± 0.1%, p_{E2} = 6.8 ± 0.7%. Model 1 is in agreement with the population shift observed in Abl^{H415P} in which E_2 is eliminated and the population of E_1 is doubled.

Measurement of the populations of the Abl conformational states

Populations of the active (A) and the two inactive states (I_1 and I_2) in the several Abl variants studied in this work were measured by fitting the CEST data as described above and by integrating the resonances corresponding to each state using ^1H-^{13}C–correlated spectra. Because the interconversion of states is slow on the NMR time scale, a distinct set of resonances are present in a 2D correlated spectrum, provided that its population is higher than ~5%. This latter approach was used for Abl variants that did not remain stable over the period of several days typically needed for recording the CEST data. It was successfully used to measure the populations in Abl^{FK}, which gave very poor CEST data because of low solubility and stability. Two variants were produced and tested for Abl^{FK}: one that encompassed residues 1 to 534 and one that encompassed residues 83 to 534. Both include the full regulatory module consisting of the SH3-SH2 tandem. Given that both variants gave very similar results, and because of the much higher stability of the 83-to-534 fragment, the latter one was primarily used. Changes in populations as a result of a mutation were expressed in free energy terms as $\Delta\Delta G = -RT(\ln R_W - \ln R_M)$ in kcal mol^{-1}, where R_W and R_M are the ratios of the populations of the two states in the wild-type and the mutant, respectively.

Structure determination

Structures were calculated from experimental methyl-methyl, methyl-aromatic, aromatic-aromatic, methyl-NH, and aromatic-NH NOEs (table S1). Dihedral angles and hydrogen-bond restraints were derived from NOE patterns. Complete NOE spectroscopy (NOESY) spectra assignment conducted iteratively from sparse restraint sets (*98*) using the automated NOESY cross-peak assignment protocol and dihedral angle dynamics simulated annealing within CYANA 3.98 (*99*). NOE peak assignments were manually curated and refined with several CYANA runs. The lowest target function CYANA-derived structure ensemble was subjected to an all-atom restrained molecular dynamics energy refinement in water bath using CNS (*100*). The active conformation is the predominant (~90%) one populated by Abl in solution, and hence its structure was determined using standard approaches. By contrast, I_1 and I_2 are marginally stable, with either one having a population of 6%. As described in the main text, to structurally characterize these states by NMR, we sought to increase their population by amino acid substitutions to shift the equilibrium toward these states without eliciting other structural changes. Because we had determined the chemical shifts of the I_1 and I_2 states by CEST, it was straightforward to assess the effect that the amino acid substitutions had on their structure. The M309L/H415P double substitution was found to markedly stabilize the I_1 state with a population of ~50% (fig. S5). NOESY data were then collected using this sample. Detailed chemical shift analyses showed that the structure of the I_1 state is similar to the one of Abl in complex with the PD173955. Especially in protein parts that are remote to the catalytic site, the A-loop and the αC helix, the two structures were essentially identical. For this reason, we used restraints from the Abl-PD173955 for areas that do not change in Abl I_1, whereas we determined de novo the structure of Abl I_1 in all others using the NMR restraints collected on the $Abl^{M309L/H415P}$ variant. To determine the structure of the Abl I_2 state, we followed a similar approach. We initially collected a large number of NOEs on the $Abl^{M309L/T408Y}$ variant, which populates the I_2 state at ~85%. Using the preliminary structural data, we gained more insight into the effect of several other mutants, and some of them were observed to further increase the population of the I_2 state, but even more importantly, they exhibited much sharper peaks and were more stable. The $Abl^{G269E/M309L/T408Y}$ variant (also referred to as the I2M variant in the main text) increases the I_2 population to 90% and was amenable to higher-resolution NMR characterization.

Abl phosphorylation by Hck

To prepare Abl specifically phosphorylated at the Tyr^{412} residue, we used Hck kinase. Two μM Hck was autophosphorylated in 50 mM Tris buffer (pH 7.5) containing 2 mM ATP, 10 mM $MgCl_2$, 50 mM NaCl, and 3.0 mM BME for 1.5 hours at 22°C. Abl (~5 μM) was incubated with 250 nM of autophosphorylated Hck for 14 to 16 hours at 15°C, followed by extensive dialysis against 50 mM Tris buffer (pH 7.5), 50 mM NaCl, and 3.0 mM BME to remove ATP, ADP, and $MgCl_2$ and ion exchange purification (Source 15Q, GE Healthcare) to separate the phosphorylated Abl from any traces of unphosphorylated protein. Abl phosphorylation was monitored by Western blot analysis with an antibody specific to Abl $pTyr^{412}$ (2865S, Cell Signaling Technology). Western blot indicated that phosphorylation of Abl by Hck led to the phosphorylation of Tyr^{412}, and mass spectrometry and NMR spectra confirmed that the samples were homogeneous.

Kinase assays

The kinase assays were conducted in 50 mM Tris (pH 7.5), 0.5 mM $MgCl_2$, 100 mM KCl, and 3.0 mM BME at room temperature with CrkII as substrate. Ten μM CrkII was incubated with 0.2 μM Abl in the reaction buffer, with the addition of ATP to a final concentration of 0.1 mM to initiate the reaction. Reactions were stopped at 0, 30, 60, 120, 180, 300, and 600 s by adding SDS-containing loading buffer. Proteins were resolved in a 4 to 20% SDS–PAGE (polyacrylamide gel electrophoresis) gel (Bio-Rad Laboratories) and probed by Western blot with antibody to CrkII $pTyr^{221}$ (3491S, Cell Signaling Technology). All blots were loaded with SuperSignal West Pico Plus Chemiluminescent Substrate (34577, Thermo Scientific), visualized using an Amersham Imager 600 (GE Life Sciences), and quantified with ImageJ. Activities of Abl proteins on CrkII are shown as the *n*-fold activation above the activity at 0 s for each time point.

REFERENCES AND NOTES

1. M. Huse, J. Kuriyan, The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002). doi: 10.1016/S0092-8674(02)00741-9; pmid: 12015977
2. R. Bayliss, T. Haq, S. Yeoh, The Ys and wherefores of protein kinase autoinhibition. *Biochim. Biophys. Acta* **1854**, 1586–1594 (2015). doi: 10.1016/j.bbapap.2015.04.025; pmid: 25936518
3. N. H. Shah, J. F. Amacher, L. M. Nocka, J. Kuriyan, The Src module: An ancient scaffold in the evolution of cytoplasmic tyrosine kinases. *Crit. Rev. Biochem. Mol. Biol.* **53**, 535–563 (2018). doi: 10.1080/10409238.2018.1495173; pmid: 30183386
4. P. Lahiry, A. Torkamani, N. J. Schork, R. A. Hegele, Kinase mutations in human disease: Interpreting genotype-phenotype relationships. *Nat. Rev. Genet.* **11**, 60–74 (2010). doi: 10.1038/nrg2707; pmid: 20019687
5. E. D. Fleuren, L. Zhang, J. Wu, R. J. Daly, The kinome 'at large' in cancer. *Nat. Rev. Cancer* **16**, 83–98 (2016). doi: 10.1038/nrc.2015.18; pmid: 26822576
6. J. Zhang, P. L. Yang, N. S. Gray, Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009). doi: 10.1038/nrc2559; pmid: 19104514
7. S. Müller, A. Chaikuad, N. S. Gray, S. Knapp, The ins and outs of selective kinase inhibitor development. *Nat. Chem. Biol.* **11**, 818–821 (2015). doi: 10.1038/nchembio.1938; pmid: 26485069
8. P. A. Futreal *et al.*, A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004). doi: 10.1038/nrc1299; pmid: 14993899
9. S. S. Taylor, A. P. Kornev, Protein kinases: Evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011). doi: 10.1016/j.tibs.2010.09.006; pmid: 20971646
10. M. Tong, M. A. Seeliger, Targeting conformational plasticity of protein kinases. *ACS Chem. Biol.* **10**, 190–200 (2015). doi: 10.1021/cb500870a; pmid: 25486330
11. Y. Xiao, J. C. Liddle, A. Pardi, N. G. Ahn, Dynamics of protein kinases: Insights from nuclear magnetic resonance. *Acc. Chem. Res.* **48**, 1106–1114 (2015). doi: 10.1021/acs.accounts.5b00001; pmid: 25803188

12. S. Cyphers, E. F. Ruff, J. M. Behr, J. D. Chodera, N. M. Levinson, A water-mediated allosteric network governs activation of Aurora kinase A. *Nat. Chem. Biol.* **13**, 402–408 (2017). doi: 10.1038/nchembio.2296; pmid: 28166210
13. J. A. H. Gilburt *et al*., Dynamic equilibrium of the Aurora A kinase activation loop revealed by single-molecule spectroscopy. *Angew. Chem. Int. Ed.* **56**, 11409–11414 (2017). doi: 10.1002/anie.201704654; pmid: 28700101
14. J. Kim *et al*., A dynamic hydrophobic core orchestrates allostery in protein kinases. *Sci. Adv.* **3**, e1600663 (2017). doi: 10.1126/sciadv.1600663; pmid: 28435869
15. Y. Meng *et al*., Predicting the conformational variability of Abl tyrosine kinase using molecular dynamics simulations and Markov state models. *J. Chem. Theory Comput.* **14**, 2721–2732 (2018). doi: 10.1021/acs.jctc.7b01170; pmid: 29474075
16. P. M. Ung, R. Rahman, A. Schlessinger, Redefining the protein kinase conformational space with machine learning. *Cell Chem. Biol.* **25**, 916–924.e2 (2018). doi: 10.1016/j.chembiol.2018.05.002; pmid: 29861272
17. N. M. Levinson *et al*., A Src-like inactive conformation in the Abl tyrosine kinase domain. *PLOS Biol.* **4**, e144 (2006). doi: 10.1371/journal.pbio.0040144; pmid: 16640460
18. N. Jura *et al*., Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol. Cell* **42**, 9–22 (2011). doi: 10.1016/j.molcel.2011.03.004; pmid: 21474065
19. S. S. Taylor, H. S. Meharena, A. P. Kornev, Evolution of a dynamic molecular switch. *IUBMB Life* **71**, 672–684 (2019). doi: 10.1002/iub.2059; pmid: 31059206
20. J. Colicelli, ABL tyrosine kinases: Evolution of function, regulation, and specificity. *Sci. Signal.* **3**, re6 (2010). doi: 10.1126/scisignal.3139re6; pmid: 20841568
21. E. K. Greuber, P. Smith-Pearson, J. Wang, A. M. Pendergast, Role of ABL family kinases in cancer: From leukaemia to solid tumours. *Nat. Rev. Cancer* **13**, 559–571 (2013). doi: 10.1038/nrc3563; pmid: 23842646
22. S. Panjarian, R. E. Iacob, S. Chen, J. R. Engen, T. E. Smithgall, Structure and dynamic regulation of Abl kinases. *J. Biol. Chem.* **288**, 5443–5450 (2013). doi: 10.1074/jbc.R112.438382; pmid: 23316053
23. B. J. Druker *et al*., Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N. Engl. J. Med.* **355**, 2408–2417 (2006). doi: 10.1056/NEJMoa062867; pmid: 17151364
24. M. E. Gorre *et al*., Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**, 876–880 (2001). doi: 10.1126/science.1062538; pmid: 11423618
25. S. Roumiantsev *et al*., Clinical resistance to the kinase inhibitor STI-571 in chronic myeloid leukemia by mutation of Tyr-253 in the Abl kinase domain P-loop. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10700–10705 (2002). doi: 10.1073/pnas.162140299; pmid: 12149456
26. N. P. Shah *et al*., Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**, 117–125 (2002). doi: 10.1016/S1535-6108(02)00096-X; pmid: 12204532
27. M. Azam, R. R. Latek, G. Q. Daley, Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831–843 (2003). doi: 10.1016/S0092-8674(03)00190-9; pmid: 12654249
28. W. W. Chan *et al*., Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036. *Cancer Cell* **19**, 556–568 (2011). doi: 10.1016/j.ccr.2011.03.003; pmid: 21481795
29. S. Wiesner *et al*., A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. *EMBO J.* **25**, 4686–4696 (2006). doi: 10.1038/sj.emboj.7601315; pmid: 16977320
30. N. Amatya *et al*., Lipid-targeting pleckstrin homology domain turns its autoinhibitory face toward the TEC kinases. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21539–21544 (2019). doi: 10.1073/pnas.1907566116; pmid: 31591208
31. Y. Xiao *et al*., Phosphorylation releases constraints to domain motion in ERK2. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2506–2511 (2014). doi: 10.1073/pnas.1318899111; pmid: 24550275
32. L. Skora, J. Mestan, D. Fabbro, W. Jahnke, S. Grzesiek, NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4437–E4445 (2013). doi: 10.1073/pnas.1314712110; pmid: 24191057
33. Y. Tokunaga, K. Takeuchi, H. Takahashi, I. Shimada, Allosteric enhancement of MAP kinase p38α's activity and substrate selectivity by docking interactions. *Nat. Struct. Mol. Biol.* **21**, 704–711 (2014). doi: 10.1038/nsmb.2861; pmid: 25038803
34. P. C. Aoto, B. T. Martin, P. E. Wright, NMR characterization of information flow and allosteric communities in the MAP kinase p38γ. *Sci. Rep.* **6**, 28655 (2016). doi: 10.1038/srep28655; pmid: 27353957
35. H. Chen *et al*., Elucidation of a four-site allosteric network in fibroblast growth factor receptor tyrosine kinases. *eLife* **6**, e21137 (2017). doi: 10.7554/eLife.21137; pmid: 28166054
36. A. Piserchio *et al*., Local destabilization, rigid body, and fuzzy docking facilitate the phosphorylation of the transcription factor Ets-1 by the mitogen-activated protein kinase ERK2. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6287–E6296 (2017). doi: 10.1073/pnas.1702973114; pmid: 28716922
37. G. S. Kumar *et al*., Dynamic activation and regulation of the mitogen-activated protein kinase p38. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4655–4660 (2018). doi: 10.1073/pnas.1721441115; pmid: 29666261
38. B. Nolen, S. Taylor, G. Ghosh, Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675 (2004). doi: 10.1016/j.molcel.2004.08.024; pmid: 15350212
39. T. R. Alderson, L. E. Kay, Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.* **60**, 39–49 (2020). doi: 10.1016/j.sbi.2019.10.008; pmid: 31835059
40. E. Rennella, R. Huang, A. Velyvis, L. E. Kay, $^{13}CHD_2$-CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins. *J. Biomol. NMR* **63**, 187–199 (2015). doi: 10.1007/s10858-015-9974-z; pmid: 26271302
41. P. Vallurupalli, A. Sekhar, T. Yuwen, L. E. Kay, Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: A primer. *J. Biomol. NMR* **67**, 243–271 (2017). doi: 10.1007/s10858-017-0099-4; pmid: 28317074
42. B. B. Brasher, R. A. Van Etten, c-Abl has high intrinsic tyrosine kinase activity that is stimulated by mutation of the Src homology 3 domain and by autophosphorylation at two distinct regulatory tyrosines. *J. Biol. Chem.* **275**, 35631–35637 (2000). doi: 10.1074/jbc.M005401200; pmid: 10964922
43. M. A. Seeliger *et al*., Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. *Cancer Res.* **69**, 2384–2392 (2009). doi: 10.1158/0008-5472.CAN-08-3953; pmid: 19276351
44. N. Vajpai *et al*., Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *J. Biol. Chem.* **283**, 18292–18302 (2008). doi: 10.1074/jbc.M801337200; pmid: 18434310
45. J. S. Tokarski *et al*., The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res.* **66**, 5790–5797 (2006). doi: 10.1158/0008-5472.CAN-05-4187; pmid: 16740718
46. B. Nagar *et al*., Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **62**, 4236–4243 (2002). pmid: 12154025
47. N. M. Levinson, S. G. Boxer, Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLOS ONE* **7**, e29828 (2012). doi: 10.1371/journal.pone.0029828; pmid: 22493660
48. T. Pemovska *et al*., Axitinib effectively inhibits BCR-ABL1 (T315I) with a distinct binding conformation. *Nature* **519**, 102–105 (2015). doi: 10.1038/nature14119; pmid: 25686603
49. E. Weisberg *et al*., Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **7**, 129–141 (2005). doi: 10.1016/j.ccr.2005.01.007; pmid: 15710326
50. T. Zhou *et al*., Structural mechanism of the Pan-BCR-ABL inhibitor ponatinib (AP24534): Lessons for overcoming kinase inhibitor resistance. *Chem. Biol. Drug Des.* **77**, 1–11 (2011). doi: 10.1111/j.1747-0285.2010.01054.x; pmid: 21118377
51. M. A. Young *et al*., Structure of the kinase domain of an imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680. *Cancer Res.* **66**, 1007–1014 (2006). doi: 10.1158/0008-5472.CAN-05-2788; pmid: 16424036
52. B. Nagar *et al*., Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859–871 (2003). doi: 10.1016/S0092-8674(03)00194-6; pmid: 12654251
53. G. Bouvignies *et al*., Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011). doi: 10.1038/nature10349; pmid: 21857680
54. L. Chen *et al*., Structural instability tuning as a regulatory mechanism in protein-protein interactions. *Mol. Cell* **44**, 734–744 (2011). doi: 10.1016/j.molcel.2011.09.022; pmid: 22152477
55. S. R. Tzeng, C. G. Kalodimos, Allosteric inhibition through suppression of transient conformational states. *Nat. Chem. Biol.* **9**, 462–465 (2013). doi: 10.1038/nchembio.1250; pmid: 23644478
56. Y.-L. Lin, Y. Meng, W. Jiang, B. Roux, Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1664–1669 (2013). doi: 10.1073/pnas.1214330110; pmid: 23319661
57. C. Wilson *et al*., Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882–886 (2015). doi: 10.1126/science.aaa1823; pmid: 25700521
58. D. W. Sherbenou *et al*., BCR-ABL SH3-SH2 domain mutations in chronic myeloid leukemia patients on imatinib. *Blood* **116**, 3278–3285 (2010). doi: 10.1182/blood-2008-10-183665; pmid: 20519627
59. M. I. Davis *et al*., Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–1051 (2011). doi: 10.1038/nbt.1990; pmid: 22037378
60. S. Branford *et al*., Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis. *Blood* **102**, 276–283 (2003). doi: 10.1182/blood-2002-09-2896; pmid: 12623848
61. A. P. Kornev, S. S. Taylor, L. F. Ten Eyck, A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14377–14382 (2008). doi: 10.1073/pnas.0807988105; pmid: 18787129
62. M. Azam, M. A. Seeliger, N. S. Gray, J. Kuriyan, G. Q. Daley, Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat. Struct. Mol. Biol.* **15**, 1109–1118 (2008). doi: 10.1038/nsmb.1486; pmid: 18794843
63. J. Hu *et al*., Kinase regulation by hydrophobic spine assembly in cancer. *Mol. Cell. Biol.* **35**, 264–276 (2015). doi: 10.1128/MCB.00943-14; pmid: 25348715
64. A. P. Kornev, S. S. Taylor, Dynamics-driven allostery in protein kinases. *Trends Biochem. Sci.* **40**, 628–647 (2015). doi: 10.1016/j.tibs.2015.09.002; pmid: 26481499
65. H. S. Meharena *et al*., Deciphering the structural basis of eukaryotic protein kinase regulation. *PLOS Biol.* **11**, e1001680 (2013). doi: 10.1371/journal.pbio.1001680; pmid: 24143133
66. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
67. A. C. Bishop *et al*., A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000). doi: 10.1038/35030148; pmid: 11014197
68. S. B. Hari, E. A. Merritt, D. J. Maly, Sequence determinants of a specific inactive protein kinase conformation. *Chem. Biol.* **20**, 806–815 (2013). doi: 10.1016/j.chembiol.2013.05.005; pmid: 23790491
69. R. E. Iacob *et al*., Conformational disturbance in Abl kinase upon mutation and deregulation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1386–1391 (2009). doi: 10.1073/pnas.0811912106; pmid: 19164531
70. D. W. Bell *et al*., Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat. Genet.* **37**, 1315–1316 (2005). doi: 10.1038/ng1671; pmid: 16258541
71. B. J. Lee, N. P. Shah, Identification and characterization of activating ABL1 1b kinase mutations: Impact on sensitivity to ATP-competitive and allosteric ABL1 inhibitors. *Leukemia* **31**, 1096–1107 (2017). doi: 10.1038/leu.2016.353; pmid: 27890928
72. C. Greenman *et al*., Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007). doi: 10.1038/nature05610; pmid: 17344846
73. J. A. Adams, Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **101**, 2271–2290 (2001). doi: 10.1021/cr000230w; pmid: 11749373
74. N. Kannan, A. F. Neuwald, Did protein kinase regulatory mechanisms evolve through elaboration of a simple

structural component? *J. Mol. Biol.* **351**, 956–972 (2005). doi: 10.1016/j.jmb.2005.06.057; pmid: 16051269
75. M. Bukhtiyarova, M. Karpusas, K. Northrop, H. V. M. Namboodiri, E. B. Springman, Mutagenesis of p38α MAP kinase establishes key roles of Phe169 in function and structural dynamics and reveals a novel DFG-OUT state. *Biochemistry* **46**, 5687–5696 (2007). doi: 10.1021/bi0622221; pmid: 17441692
76. S. H. Schmidt *et al*., The dynamic switch mechanism that leads to activation of LRRK2 is embedded in the DFGψ motif in the kinase domain. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14979–14988 (2019). doi: 10.1073/pnas.1900289116; pmid: 31292254
77. T. Saleh, P. Rossi, C. G. Kalodimos, Atomic view of the energy landscape in the allosteric regulation of Abl kinase. *Nat. Struct. Mol. Biol.* **24**, 893–901 (2017). doi: 10.1038/nsmb.3470; pmid: 28945248
78. J. Zhang *et al*., Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501–506 (2010). doi: 10.1038/nature08675; pmid: 20072125
79. A. A. Wylie *et al*., The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature* **543**, 733–737 (2017). doi: 10.1038/nature21702; pmid: 28329763
80. F. E. Kwarcinski *et al*., Conformation-selective analogues of dasatinib reveal insight into kinase inhibitor binding and selectivity. *ACS Chem. Biol.* **11**, 1296–1304 (2016). doi: 10.1021/acschembio.5b01018; pmid: 26895387
81. B. Nagar *et al*., Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. *Mol. Cell* **21**, 787–798 (2006). doi: 10.1016/j.molcel.2006.01.035; pmid: 16543148
82. P. Filippakopoulos *et al*., Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. *Cell* **134**, 793–803 (2008). doi: 10.1016/j.cell.2008.07.047; pmid: 18775312
83. S. Lorenz, P. Deng, O. Hantschel, G. Superti-Furga, J. Kuriyan, Crystal structure of an SH2-kinase construct of c-Abl and effect of the SH2 domain on kinase activity. *Biochem. J.* **468**, 283–291 (2015). doi: 10.1042/BJ20141492; pmid: 25779001
84. F. Grebien *et al*., Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. *Cell* **147**, 306–319 (2011). doi: 10.1016/j.cell.2011.08.046; pmid: 22000011
85. B. H. Honig, W. L. Hubbell, Stability of "salt bridges" in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5412–5416 (1984). doi: 10.1073/pnas.81.17.5412; pmid: 6591197
86. Y. Shan *et al*., A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 139–144 (2009). doi: 10.1073/pnas.0811223106; pmid: 19109437
87. Y. Meng, Y. L. Lin, B. Roux, Computational study of the "DFG-flip" conformational transition in c-Abl and c-Src tyrosine kinases. *J. Phys. Chem. B* **119**, 1443–1456 (2015). doi: 10.1021/jp511792a; pmid: 25548962
88. S. R. Hubbard, L. Wei, W. A. Hendrickson, Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754 (1994). doi: 10.1038/372746a0; pmid: 7997262
89. M. A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010). doi: 10.1016/j.cell.2010.06.011; pmid: 20602996
90. W. Jankowski *et al*., Domain organization differences explain Bcr-Abl's preference for CrkL over CrkII. *Nat. Chem. Biol.* **8**, 590–596 (2012). doi: 10.1038/nchembio.954; pmid: 22581121
91. P. Rossi, Y. R. Monneau, Y. Xia, Y. Ishida, C. G. Kalodimos, Toolkit for NMR studies of methyl-labeled proteins. *Methods Enzymol.* **614**, 107–142 (2019). doi: 10.1016/bs.mie.2018.08.036; pmid: 30611422
92. Y. R. Monneau *et al*., Exploiting *E. coli* auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *J. Biomol. NMR* **65**, 99–108 (2016). doi: 10.1007/s10858-016-0041-1; pmid: 27255761
93. F. Delaglio *et al*., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995). doi: 10.1007/BF00197809; pmid: 8520220
94. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015). doi: 10.1093/bioinformatics/btu830; pmid: 25505092
95. P. Rossi, Y. Xia, N. Khanra, G. Veglia, C. G. Kalodimos, ^{15}N and ^{13}C- SOFAST-HMQC editing enhances 3D-NOESY sensitivity in highly deuterated, selectively [^{1}H,^{13}C]-labeled proteins. *J. Biomol. NMR* **66**, 259–271 (2016). doi: 10.1007/s10858-016-0074-5; pmid: 27878649
96. Y. R. Monneau *et al*., Automatic methyl assignment in large proteins by the MAGIC algorithm. *J. Biomol. NMR* **69**, 215–227 (2017). doi: 10.1007/s10858-017-0149-y; pmid: 29098507
97. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012). doi: 10.1021/ja3001419; pmid: 22554188
98. O. F. Lange *et al*., Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10873–10878 (2012). doi: 10.1073/pnas.1203013109; pmid: 22733734
99. P. Güntert, L. Buchner, Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015). doi: 10.1007/s10858-015-9924-9; pmid: 25801209
100. A. T. Brunger, Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007). doi: 10.1038/nprot.2007.406; pmid: 18007608

ACKNOWLEDGMENTS

We thank Z. Xia for help with molecular cloning, Y. Xia for help with the NMR experiments, Z. Luo for producing the animations, G. Bouvignies for providing the ChemEx software, L. Kay and R. Anderson for discussions, and B. Young for the synthesis of the labeled Phe and Tyr. NMR spectra were acquired at the St. Jude Biomolecular NMR Spectroscopy Center. **Funding:** This work was supported by U.S. National Institutes of Health grant R35 GM122462 (C.G.K.) and ALSAC. **Author contributions:** T.X. and C.G.K. conceived the study and designed the experimental approach. T.X. prepared all protein samples, determined the Abl structures, and performed the activity assays. T.S. assisted with the kinase assays, and P.R. assisted with the structure determination. T.X. and C.G.K. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic coordinates and NMR chemical shifts have been deposited in the Protein Data Bank (PDB) and Biological Magnetic Resonance Data Bank (BMRB), respectively, under the following accession codes: 6XR6 and 30770 for the active state of Abl, 6XR7 and 30771 for the inactive state 1 (I_1) of Abl, and 6XRG and 30772 for the inactive state 2 (I_2) of Abl. All other data are available in the main text or the supplementary materials. Requests for materials should be addressed to the corresponding author.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/eabc2754/suppl/DC1
Figs. S1 to S12
Table S1
References
MDAR Reproducibility Checklist
Movies S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

16 April 2020; accepted 12 August 2020
Published online 1 October 2020
10.1126/science.abc2754

RESEARCH ARTICLE SUMMARY

METALLURGY

The heterogeneity of persistent slip band nucleation and evolution in metals at the micrometer scale

Steven Lavenstein, Yejun Gu, Dylan Madisetti, Jaafar A. El-Awady*

INTRODUCTION: Metals are the material of choice for many structural applications because they provide the best compromise between strength and ductility. For applications in which cyclic loading is imposed, fatigue failure plagues all metals, and mitigating it is of great importance. In ductile metals, fatigue cracks initiate as small, microstructurally short cracks that gradually grow with increasing number of loading cycles. Although many studies have been dedicated to the crack-growth stage, the transition from a crack-free to a cracked metal remains one of the most challenging topics in the study of fatigue of metal.

The nucleation of microcracks in ductile metals is a consequence of the to-and-fro motion of dislocations during cyclic loading, which leads to dislocation self-organization into long-range ordered structures. Dislocations result from irregularities in the arrangement of atoms in crystalline materials, and their motion leads to plastic deformation. Ladder dislocation structures, more commonly referred to as persistent slip bands (PSBs), are perhaps the most consequential defect structures with regard to fatigue crack initiation. PSBs take the form of regularly spaced, dislocation-dense walls constructed of edge dislocation dipoles with dislocation-sparse channels separating them in a structure resembling a ladder.

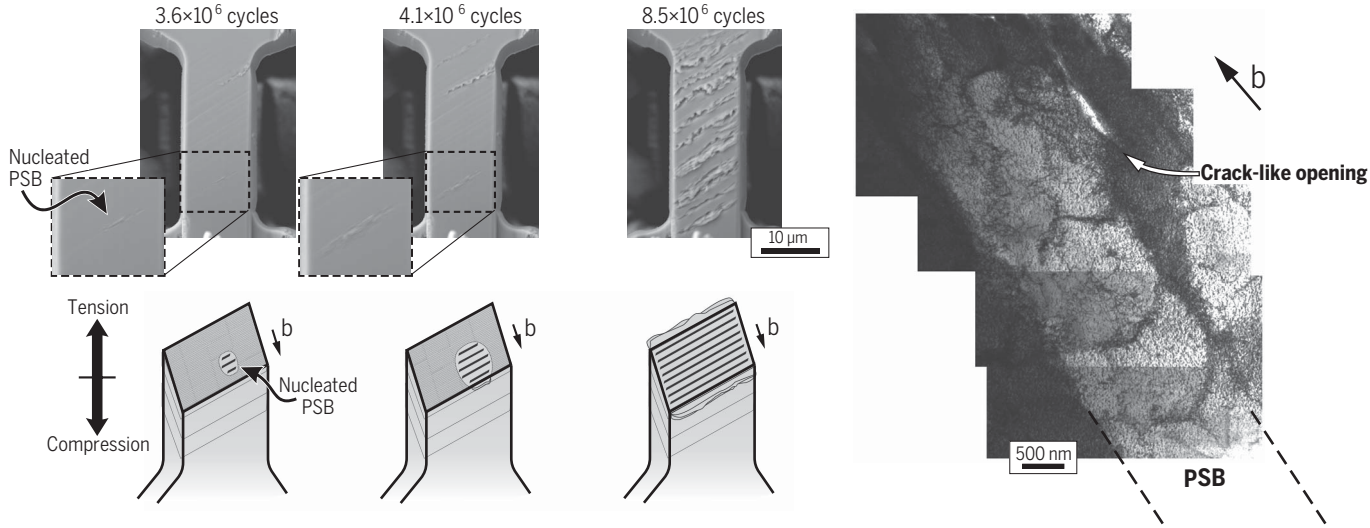
RATIONALE: Our aim is to provide in situ observations and characterization of the formation of PSBs in micrometer-sized Ni single crystals—a representative, model face-centered cubic metal. To do this, we designed a high-frequency microfatigue experiment that replicates the necessary conditions for PSB formation in a very confined material volume. We conducted all experiments in situ in a scanning electron microscope (SEM) on microcrystals having rectangular cross sections with nominal dimensions of 12 μm by 13 μm and gauge lengths of 27 μm. We oriented all microcrystals for single slip, with a fully reversible and symmetric tension-compression cyclic loading imposed along the $[3\bar{5}2]$ crystallographic direction at a constant shear strain amplitude below 1.5×10^{-2} and a loading frequency of 75 Hz. We generated propagation profiles of PSB surface slip markings from the in situ observations, and the dislocation structure postmortem was characterized with transmission electron microscopy (TEM).

RESULTS: We found that PSBs nucleate locally within the microcrystal volume and then propagate gradually until they span the entire slip region. A relatively large number of cycles ($>10^6$) was necessary to nucleate PSBs in microcrystals compared with bulk scale, and correspondingly, extreme fatigue lifetimes were exhibited at the micrometer scale. The PSB surface slip markings also seem to have an inherent roughness immediately on formation; after fully propagating, the roughness of the PSB slip markings remains stable with further cyclic loading. The slip traces formed in the first ~10 cycles are also found to identify the locations where PSBs—and thus cracks—eventually form.

CONCLUSION: This gradual evolution suggests certain refinements to conventional PSB models that idealize extrusion formation as a rapid growth process. This is important for nondestructive damage quantification (i.e., failure prediction) and for the physics-based modeling community, because simulating the evolution of fatigue damage is usually too computationally expensive to perform beyond a few hundred cycles. Although the results presented in this paper focus on pure Ni, the fundamental mechanisms identified are common to many metals. Thus, the current insights provide an avenue to connect micrometer-scale deformation mechanisms with fatigue failures at the bulk scale in metals. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jelawady@jhu.edu
Cite this article as S. Lavenstein *et al.*, *Science* **370**, eabb2690 (2020). DOI: 10.1126/science.abb2690

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abb2690



High-frequency microfatigue experiments on Ni single crystals. Successive micrographs from an SEM video (top left) taken during an in situ cyclic loading experiment show the nucleation and propagation of PSB surface slip markings. The proposed dislocation organization mechanisms that explain this are also illustrated schematically (bottom left). A TEM micrograph (right) confirms the presence of the ladderlike self-organized dislocation structure, which is the hallmark of PSBs.

RESEARCH ARTICLE

METALLURGY

The heterogeneity of persistent slip band nucleation and evolution in metals at the micrometer scale

Steven Lavenstein*, Yejun Gu, Dylan Madisetti, Jaafar A. El-Awady†

Fatigue damage in metals manifests itself as irreversible dislocation motion followed by crack initiation and propagation. Characterizing the transition from a crack-free to a cracked metal remains one of the most challenging problems in fatigue. Persistent slip bands (PSBs) form in metals during cyclic loading and are one of the most important aspects of this transition. We used in situ microfatigue experiments to investigate PSB formation and evolution mechanisms, and we discovered that PSBs are prevalent at the micrometer scale. Dislocation accumulation rates at this scale are smaller than those in bulk samples, which delays PSB nucleation. Our results suggest the need to refine PSB and crack-initiation models in metals to account for gradual and heterogeneous evolution. These findings also connect micrometer-scale deformation mechanisms with fatigue failure at the bulk scale in metals.

Metals are the material of choice for many structural applications because they provide the best compromise between strength and ductility. For applications in which cyclic loading is imposed, the mitigation of fatigue damage is important. Ductile metals are preferred for such applications because they fail in a more gradual manner. This behavior is in contrast to brittle materials, which fail catastrophically and with little warning. In ductile metals, fatigue cracks initiate as small, microstructurally short cracks and gradually grow with an increasing number of loading cycles. Although many studies have been dedicated to the crack-growth stage, the transition from a crack-free metal to a cracked metal remains one of the most challenging topics in the study of fatigue of metals.

The nucleation of microcracks in ductile metals is a consequence of the to-and-fro motion of dislocations during cyclic loading, which leads to dislocation self-organization into long-range ordered structures, such as dislocation veins, cells, and/or ladder structures (*1*, *2*). Ladder dislocation structures, more commonly referred to as persistent slip bands (PSBs), are perhaps the most consequential of these structures with regard to fatigue crack initiation. PSBs take the form of regularly spaced, dislocation-dense walls constructed of edge dislocation dipoles with dislocation-sparse channels separating them in a structure resembling a ladder. PSBs are typically observed in wavy slip metals, in large grains or crystals oriented for single slip, with a symmetrically applied tension-compression loading profile, and along the slip plane with the highest resolved shear stress (*3*, *4*). The dislocation walls organize along the direction of the slip plane where the atoms are most closely packed (i.e., the Burgers vector direction) and egress from the free surface of the crystal in the direction of the Burgers vector as surface steps. Additionally, PSBs have been found to promote microcrack initiation along the interface between the PSB and the surrounding matrix at the crystal-free surface (*5*) or at the intersection of the PSBs with high-angle grain boundaries.

The precise mechanisms that lead to surface roughness and eventually to fracture along the PSB-matrix interface are a subject of debate. One of the leading theories is that vacancies form within the PSB as it is cyclically deformed, which leads to a volume increase in the PSB. This results in a build-up of stress that is eventually relieved by dislocations departing from the free surface of the crystal and leaving a pronounced extrusion. This model is commonly known as the EGM model, named after the authors who originally proposed it (*6*). Although this aspect of the PSB is generally agreed upon, much debate revolves around the subsequent steps. The EGM model proposes that if vacancies are immobile because of low temperatures, the extrusion height will stabilize once the PSB occupies the entire crystal diameter. However, statistical roughening of the PSB extrusion will still occur because the PSB carries a large amount of plastic strain, not all of which is reversible. Eventually, a microcrack will form at the PSB-matrix interface as a result of the stress concentration induced by the extrusion. If the temperatures are high enough, the EGM model predicts coalescence of vacancies through diffusion. However, this diffusion is hypothesized to be confined to the PSB-matrix interface, which results in the formation of thin, crack-like intrusions alongside the PSB. In a modification of this model, Polák has proposed that the diffusion of vacancies occurs throughout the PSB volume (*7*). One consequence of this is that it may lead to an initial tonguelike extrusion profile where extrusions and intrusions alternate along the surface. The Polák model regards these intrusions—one on each side of a PSB—to be a dominant factor in crack initiation. Proponents of the EGM model, however, believe that the extrusion stress concentration is more prevalent in fatigue crack initiation.

Generally, PSBs are classified as long-range features, and although they have been commonly observed in bulk polycrystalline metals with grain sizes as small as tens of micrometers, they have never been reported in single crystals with sizes smaller than a few hundred micrometers (*8*–*10*). Additionally, only recent advances in in situ, small-scale testing have enabled cyclic loading of micrometer-scale materials to millions of cycles (*11*–*15*). Such in situ methods are needed to shed light on the deformation mechanisms involved, because most bulk-scale fatigue studies have primarily focused on post-mortem characterization of the formation of PSBs, with very few exceptions (*16*). Accordingly, our aim is to provide in situ observations and characterization of the formation of PSBs in a micrometer-sized, representative model face-centered cubic (fcc) metal, single crystal Ni.

We conducted all experiments in situ in a scanning electron microscope (SEM) on microcrystals having rectangular cross sections with nominal dimensions of 12 μm by 13 μm and gauge lengths of 27 μm. We oriented all microcrystals for single slip, with a fully reversible and symmetric tension-compression cyclic loading imposed along the $[3\bar{5}2]$ crystallographic direction at a constant shear strain amplitude and a constant frequency of 75 Hz (*17*).

Evolution of surface morphology during cyclic loading

We collected sequential SEM micrographs of the evolution of the surface morphology on the edge plane (fig. S1) of a Ni microcrystal that we cyclically loaded at a constant shear strain amplitude of 6.3×10^{-3} (Fig. 1 and movie S1). We also measured the dynamic stiffness as a function of the number of cycles for the same microcrystal (Fig. 1Q and movie S1). Our observations from this experiment are common for all experiments conducted below shear strain amplitudes of ~1.5×10^{-2}. The shear strain amplitude range for obtaining PSBs in bulk Ni is ~10^{-3} to 10^{-2} (*18*). We observed that faint slip traces form within the first 8 loading cycles (Fig. 1B). Subsequently, the surface morphology remained relatively unchanged up to 2.5×10^6 cycles, whereas the dynamic stiffness

Department of Mechanical Engineering, Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA.
*Present address: Intelligent Automation, Inc., Rockville, MD 20855, USA.
†Corresponding author. Email: jelawady@jhu.edu

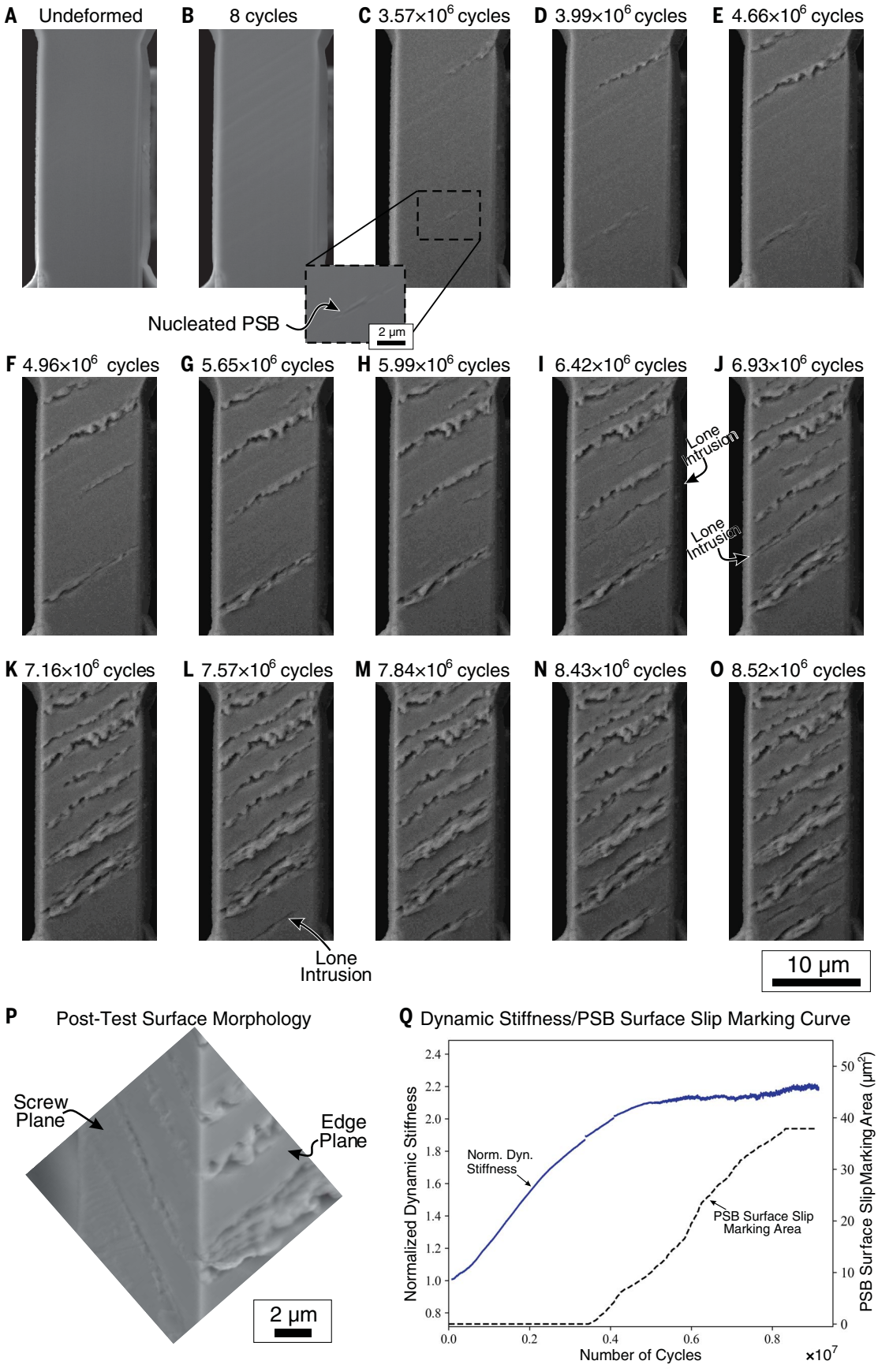


Fig. 1. Ni high-frequency microfatigue experimental results. (**A** to **O**) SEM micrographs of the evolving surface morphology on the edge plane of a Ni microcrystal cyclically loaded at a shear strain amplitude of 6.3×10^{-3}. The microcrystal has a rectangular cross section with dimensions of 12 μm by 10 μm and a gauge length of 26 μm. (**P**) Angled view of the surface morphology after testing showing both the edge and screw planes. (**Q**) The dynamic stiffness and the total area of four analyzed PSB surface slip markings on the edge plane versus the number of loading cycles. In these micrographs, intrusions are distinguished as dark regions on the surface of the microcrystal. Norm. Dyn. Stiffness indicates the normalized dynamic stiffness.

continuously increased during this stage, correlated to cyclic hardening (*13*). After a critical number of cycles was reached (typically >10^6 cycles), slip band markings nucleated and propagated sequentially on the edge plane along the different slip traces that had formed in the first ~10 cycles (Fig. 1C, inset).

These slip band markings did not appear along the entire slip plane at once. Instead, they first nucleated at a localized region of the slip plane and then spread in both directions, perpendicular to the primary Burgers vector on the slip plane, until the slip band extended the entire gauge width. The thickness of each slip band marking is ~1 μm. As the slip band markings sequentially nucleated with increasing number of cycles, the rate of increase in the dynamic stiffness decreased until it reached a steady state value when no new slip band markings were observed, which corresponds to cyclic saturation (*13*).

A closer examination of an angled view of the microcrystal (Fig. 1P) shows the surface morphology on both the edge and screw planes after testing. This angled view shows that the surface topography of the surface slip marking is composed of extrusions and intrusions. Furthermore, our observations indicate that surface intrusions form on the screw plane well after the formation of fully developed slip band markings on the edge plane, and they appear more gradually compared with the abrupt nucleation event observed on the edge plane.

Underlying dislocation structure

We extracted cross-sectional lift-out foils from the interior of the deformed specimens with a focused ion beam (FIB) (*17*) to characterize the underlying dislocation structure using transmission electron microscopy (TEM). We observed the dislocation structure in a foil that was extracted parallel to the screw plane from a microcrystal cyclically deformed at a shear strain amplitude of 1.1×10^{-3} for 10^4 cycles and then at an amplitude of 6.9×10^{-3} for 1.45×10^7 cycles (Fig. 1). The dislocation structure we observed exhibits characteristic ladderlike dislocation structures with regularly spaced dislocation walls separated by dislocation-sparse channels. The dislocation walls are separated along the Burgers vector direction with a mean spacing between the walls on the order of 1 μm. These observations are characteristic features of PSBs in pure bulk Ni at room temperature (*19*, *20*). The observation of multiple neighboring PSBs (Fig. 2A) also resembles macro-PSBs that have previously been observed in bulk fcc single crystals (*7*). These observations indicate that the slip markings we observed are PSBs. However, these PSBs are not as well defined as those observed in bulk-scale fatigue studies of fcc metals (*1*, *19*). The ladder rungs are not as straight or as evenly spaced. We believe this discrepancy with bulk observations is predominately an effect of crystal size.

Our SEM micrograph shows the surface morphology on the screw plane of a cyclically deformed microcrystal at a strain amplitude of 4.1×10^{-3} after 6.43×10^6 cycles (Fig. 3A). In this example, a single surface slip marking propagated through the thickness of the microcrystal. We also collected TEM micrographs of cross-sectional FIB lift-out foils that are parallel to the screw plane and taken from different locations (Fig. 3, B to F). Below the surface slip marking (Fig. 3D), we observed the early phase of development of a ladder dislocation structure, where the walls toward the interior have relatively few dislocations and are less straight compared with walls further along the PSB. Additionally, the surrounding matrix exhibits many dislocation tangles but not as dense clusters (Fig. 3B). We surmise that these dislocations are on primary and secondary slip systems. Although we also sometimes observe dislocation veins (Fig. 3, B and F), they are much more sporadic than the relatively consistent vein structures observed in bulk Ni studies, where approximately half of the matrix is occupied by veins (*18*–*20*). We observed long dislocation segments between the channels of the PSB as well as in neighboring regions (Fig. 3, C and D). We presume that these long segments are screw dislocations on the primary slip system and are entangled with dislocations on secondary slip systems. In bulk-scale experiments, screw dislocations are commonly observed gliding in the PSB channels (*21*, *22*). However, our observation that secondary dislocations entangle the long screw segments indicates that the PSB is still at an early stage of development. Notably, these entangled screw dislocations were not observed in more well-formed PSBs (Fig. 2). Dislocation tangles have also been observed in bulk fatigue at the interface of the vein and PSB structures (*23*) and have been assumed to be predominately edge multipoles on the primary and secondary slip systems. The argument for their appearance is that these tangles loosen through secondary dislocation annihilation, which frees the primary dislocations and subsequently leads to the formation of the PSB walls (*23*). On the basis of our observations (Fig. 3), it is clear that screw dislocations on the primary slip systems are also a critical component of these tangles. Three-dimensional discrete dislocation dynamics (DDD) simulations have also shown that screw dislocations on the primary slip system interact with dislocations on secondary slip systems to form the primary edge dipoles in PSB walls (*24*).

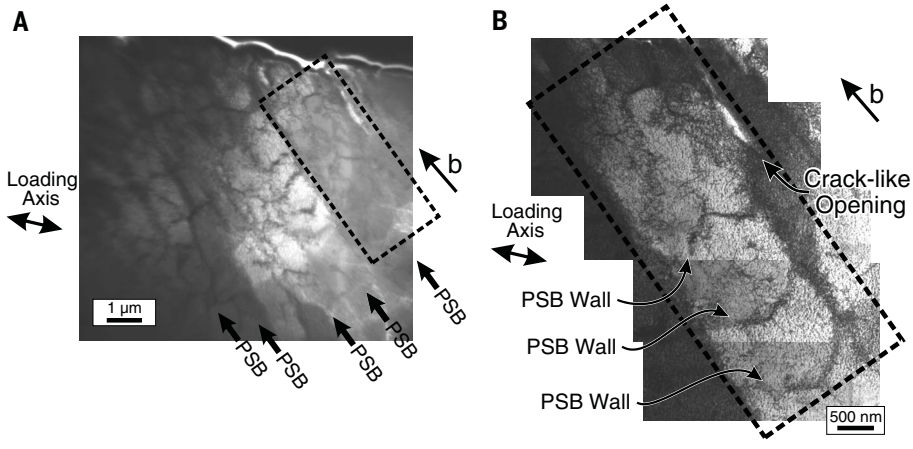


Fig. 2. Dislocation structure in a foil extracted parallel to the screw plane from a Ni microcrystal cyclically loaded at shear strain amplitudes of 1.1×10^{-3} for 10^4 cycles and then 6.9×10^{-3} for 1.45×10^7 cycles. (**A**) A macro-PSB is clear with a group of several PSBs clustered alongside one another. The b arrow indicates the Burgers vector. (**B**) A higher magnification of the dashed box in (A) showing a PSB with a crack-like opening along one side of the PSB-matrix interface.

PSB evolution profile and its embryonic width

We prepared propagation profiles of select PSB surface slip markings for Ni microcrystals cyclically loaded at different shear strain amplitudes (Fig. 4). These profiles are nonuniform and are typical of the PSB surface slip marking profiles that we observed. We determined these profiles using image analysis performed on individual frames from the in situ videos (*17*). We labeled the first cycle at which a PSB surface slip marking was detected as cycle zero. The nonuniform profiles are likely correlated with the underlying nonuniform dislocation structure that develops within the microcrystal. These profiles also indicate that the first slip markings we observed have a finite width, which we refer to as the embryonic width of the PSB surface slip marking. The embryonic widths from all of the PSB surface

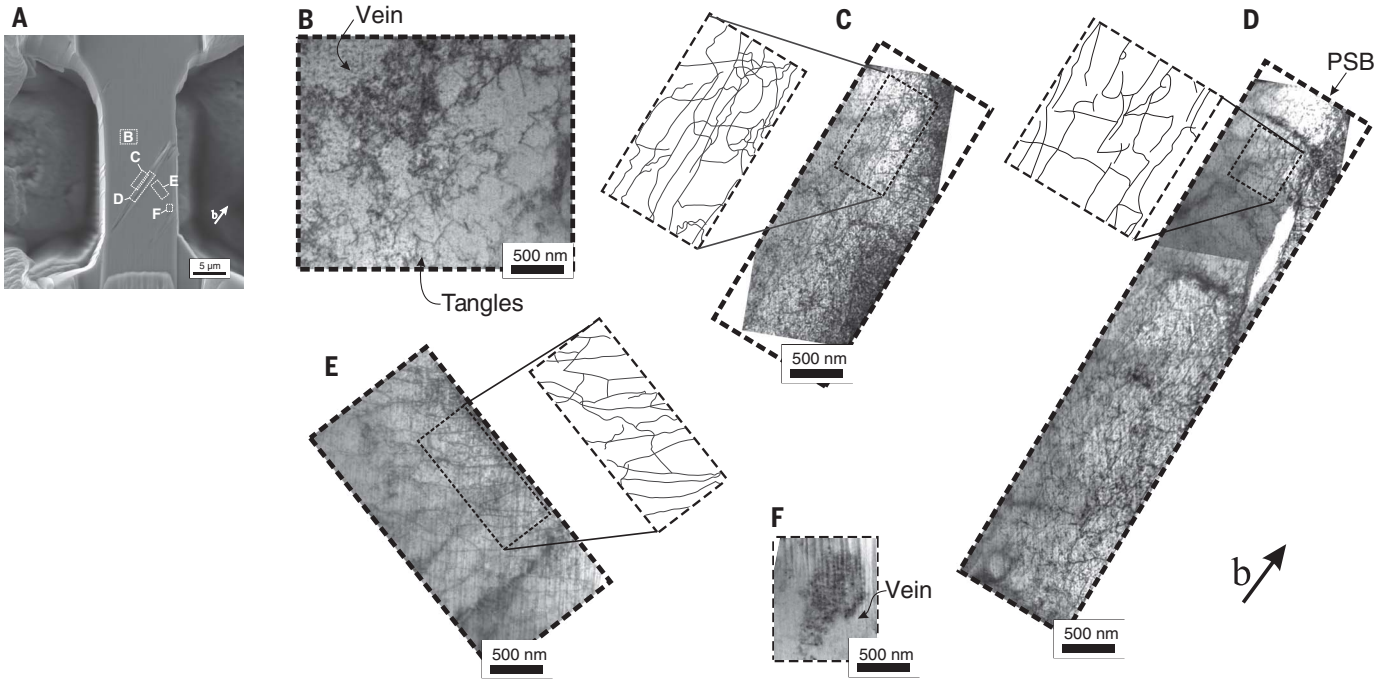


Fig. 3. SEM and TEM micrographs of a Ni microcrystal cyclically loaded at a shear strain amplitude of 4.2×10^{-3} for 6.43×10^{6} loading cycles. (**A**) SEM micrograph showing the surface morphology on the screw plane. (**B** to **F**) TEM micrographs of the dislocation structure from cross-sectional FIB lift-out foils parallel to the screw plane at different locations. Refer to the labeled boxes in (A) for the locations of (B) to (F). In (C) to (E), dislocation tangles are interpreted in a Bézier curve format (see insets) in addition to the TEM micrographs.

slip marking profiles studied here were in the range of 2.6 to 8.3 μm (*17*). The slope of the propagation profile of the PSB surface slip marking on either side of the slip band embryo width, which is on the order of 1.89 to 8.63 pm per cycle, gives us a first-order approximation of the speed at which the slip marking propagates perpendicularly to the primary Burgers vector direction. The wide variation in the embryonic widths we observed indicates that the PSB nucleus likely develops at a certain depth within the crystal volume before propagating to the surface. The variations in the PSB propagation speed is likely correlated with the nonuniformity with which the underlying dislocation structure develops in the microcrystal.

PSB nucleation and propagation model

The in situ microfatigue experiments provide a detailed view of the evolution of PSBs, from the very instant of surface nucleation to full propagation. The PSB surface slip markings first nucleate along the slip plane with an embryonic width of a few micrometers. We then observed propagation of the PSB surface slip markings in both directions perpendicular to the Burgers vectors on the primary slip plane at speeds up to a few picometers per cycle. Although PSB surface slip markings <10 μm in width have been observed in postmortem observations of bulk-scale fatigue experiments (*25*, *26*), our real-time observations suggest that these PSBs can in fact grow into larger PSBs upon further cyclic loading. This gradual evolution suggests certain refinements to the conventional PSB models that idealize extrusion formation as a rapid growth process (*27*). Namely, the statistical roughening that is surmised to happen after the PSB extrusion stabilizes may not be a necessary component of PSB models. Rather, the gradual, heterogeneous manner in which PSBs evolve throughout the slip region of a crystal may play more of a role in PSB extrusion roughness.

We developed a schematic representation of the nucleation and propagation of a PSB and the resulting surface morphology (Fig. 5A). On the first few cycles of loading, the weakest slip planes in the gauge length activate to accommodate the applied strains. The preexisting dislocations near the free surface on these weakest slip planes escape the crystal in the first few cycles, which produces faint surface slip traces. These slip traces are subsequently stable because dislocations farther into the interior of the crystal do not move far enough to escape the free surface because of the small strain amplitude imposed on the crystal. With further cyclic loading, dislocation tangles—such as those we observed (Fig. 3B)—form throughout the microcrystal, with the highest dislocation density tangles occurring around the softest slip planes. The transition from slip traces to PSB surface slip markings is directly related to the transition in the dislocation structure from tangles to PSBs. On the basis of the surface observations in this study, this appears to happen locally along a lamellar slip region adjacent to the initial slip traces. For this to happen, secondary dislocations have to unentangle and allow the primary dislocations to move freely. On the basis of DDD simulations (*24*, *28*–*31*), these primary dislocations form dipoles and cluster upon cyclic loading (Fig. 5A). This clustering forms the PSB nucleus, which egresses from the free surface in the form of surface slip markings in the direction of the primary Burgers vector.

Once a small region of the crystal exhibits a PSB with its own correspondingly small surface slip markings, the PSB will grow along the lamellar slip region (Fig. 5A). This growth has to be associated with the widening of the ladder rungs, whereas the increasing extrusion height has to be associated with more ladder rungs forming in the direction of the Burgers vector. The dislocation walls are sources for point defect production, which results in an effective local change in volume that correlates with the height and depth of the extrusion and intrusion on the surface. Once the PSB widens and fully occupies the entire crystal width, the surface slip markings are stable upon further cyclic loading. This nonhomogeneous PSB formation mechanism has also been observed in bulk-scale experiments (*17*).

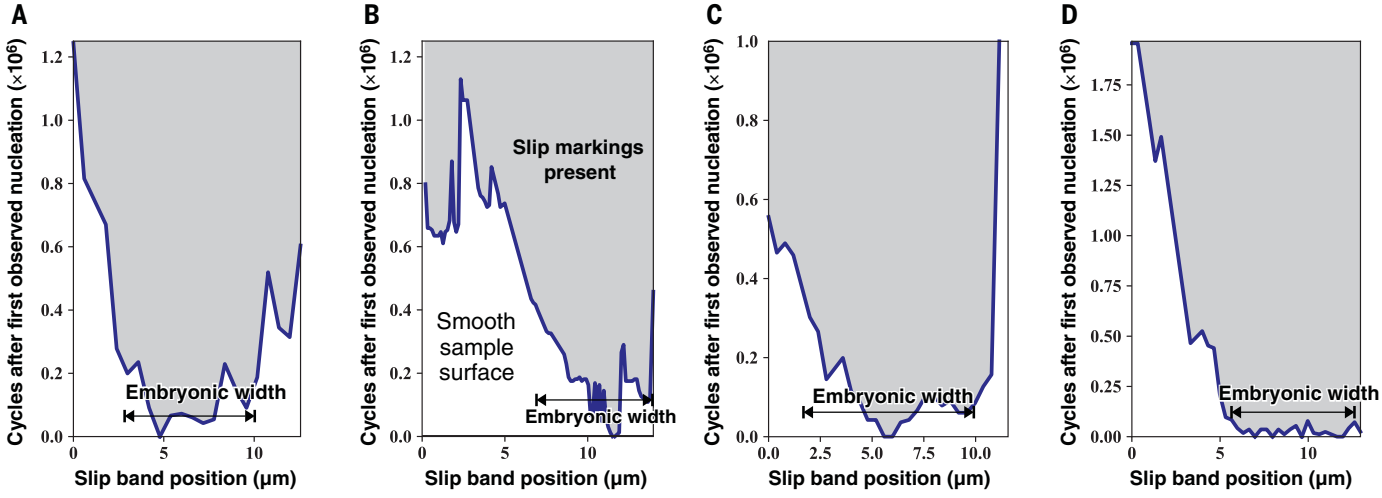


Fig. 4. Propagation profiles of different PSB surface slip markings from Ni microcrystals cyclically loaded at a shear strain amplitude of 6.3 × 10^{-3}. (**A** to **D**) The width of the PSB surface slip marking as it egresses at the free surface with a relatively flat profile is termed the embryonic width. The white regions indicate no surface marking, whereas the gray regions indicate the presence of a surface marking. The solid lines represent the transition between both regions (i.e., the PSB-cycle profile).

Comparison with previously proposed PSB models

Our real-time, in situ observations provide distinct insights with respect to PSB formation and fatigue crack initiation. Most PSB models are arguably based on the EGM model (*6*). If vacancies are considered to be immobile below a given temperature in the EGM model, the surface extrusions result from the increased vacancy volume within the PSB, and surface crack initiation is induced by the stress concentration at the interface of the PSB extrusion. The EGM model also allows for intrusion growth at the PSB-matrix interface when temperatures are high enough for vacancies to be mobile. However, regardless of the temperature, extrusions always precede any cracks or intrusions at the PSB-matrix interface in the EGM model (*32*). Our observations show many cases where extrusions precede intrusions (Fig. 1, I to K), but we found several observations where this was not the case. For example, we observed that an intrusion propagated until it spanned the entire gauge width, and only later did adjacent extrusions appear all at once (Fig. 1I). Thus, we cannot describe these features as stress-initiated cracks. A similar observation of individual intrusions preceding extrusions has also been reported in bulk fatigue (*33*).

In all samples we show here, the dominant crack-like opening appears on the acute-angle side of the PSB. We can explain this phenomenon with the EGM model by considering the asymmetric stress distribution around the PSB (i.e., the sharper notch in the vicinity of the acute angle). In a comprehensive study of PSBs, Man *et al.* have also observed this phenomenon (*34*). In addition to the notch stresses, Man *et al.* attribute this to the fact that the plastic slip at the acute-angle side of the PSB is larger and may lead to larger intrusions.

The other detail that the current results shed light on is with regard to the evolution of the surface roughness. The EGM model suggests that after the static extrusion develops, gradual roughening (i.e., statistical roughening) will occur with further cyclic loading. However, this is not observed here. Instead, the PSB extrusion seems to have an inherent surface roughness immediately on formation, and, after fully propagating, the roughness of the PSB surface slip marking remains stable with further cyclic loading.

Slip trace analysis: Early cycles versus high cycles

Our observations of the slip traces as a function of the number of fatigue cycles further elucidate the origin of PSBs and how they nucleate. The location of the slip planes that will eventually nucleate PSBs appear to be determined in the first few cycles. To show this, we compare the microcrystal surface morphology of a sample during the early stages of fatigue with the surface morphology after the PSBs have fully nucleated and propagated (Fig. 5B). The coincidence of the PSB locations with the initial slip traces is also consistent with observations from bulk Cu cyclic loading studies (*23*) and likely indicates that these planes are the weakest in the gauge length of the microcrystal. We found that the cracks and intrusions of the eventual PSBs line up with the initial slip traces, whereas the PSB extrusions appear to occupy the spaces between the initial slip traces. For any three adjacent slip traces, the PSB surface slip markings are most likely to form first between the two that are more–closely spaced together. In bulk Ni, slip traces in the first ~10 cycles have also been observed to coincide with initiated cracks with further cycling (*35*).

Influence of the specimen size

In bulk-scale experiments, high–dislocation density veins are generally observed during cyclic loading experiments, and the veins usually constitute the matrix that precedes PSB nucleation. Furthermore, once PSBs form, no further hardening occurs owing to reaching a saturation in the cyclic mechanical response. This saturation occurs after ~100 to 10^4 loading cycles at room temperature (*18*, *36*).

In contrast to bulk-scale experiments, one notable aspect of the dislocation structure we observed is the lack of high–dislocation density veins. Although we observed veinlike structures in some local regions of samples that were cyclically loaded at low shear strain amplitudes (Fig. 3, B and F), we did not observe them as consistently as in bulk fatigued Ni and Cu (*1*, *19*). Also, the veins we observed had lower dislocation densities and much smaller sizes. Instead, the matrix was mostly composed of dislocation tangles. The dislocation density distribution of these tangles was heterogeneous, with a higher density near PSBs and a lower density far from the PSBs or surrounding vein structures (Fig. 3B). We hypothesize that the loosening of these dislocation tangles is a necessary precursor for generating the free dipoles that eventually cluster into dislocation walls. Finding similar dislocation tangles within PSB channels (Fig. 3D) could mean that more dislocations could untangle on further cyclic loading. This observation is consistent with the model proposed by Tabata *et al.* (*23*) for bulk crystals, where untangling primary edge dislocations from secondary dislocations

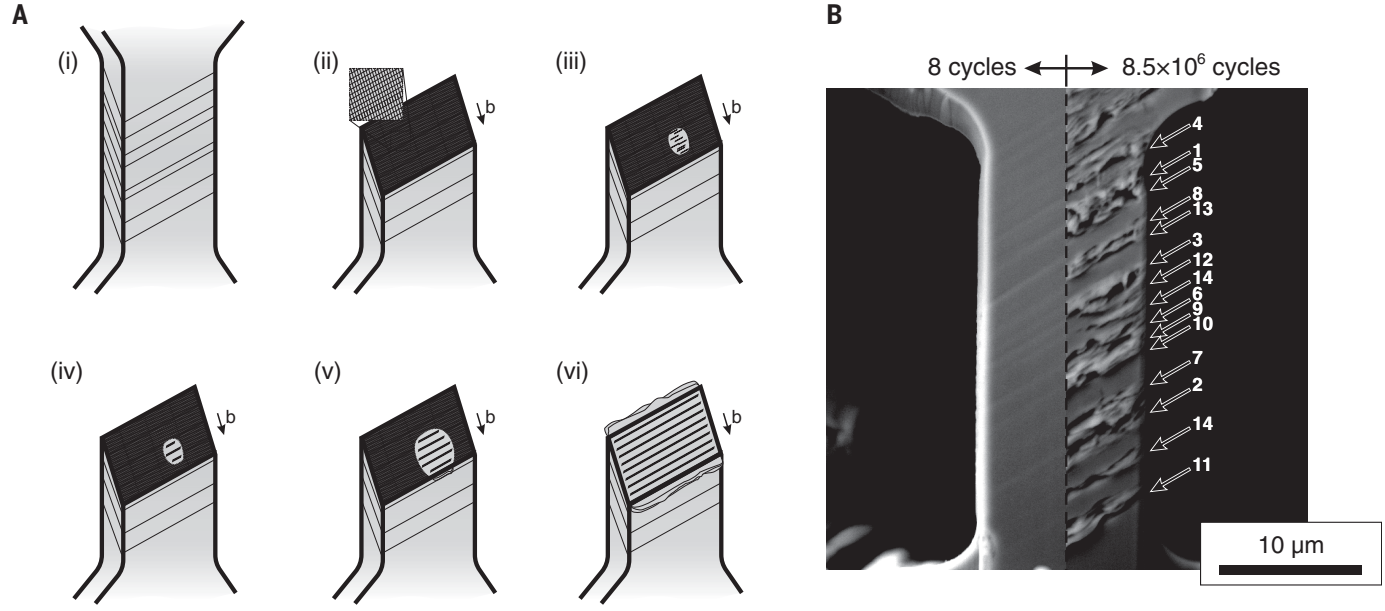


Fig. 5. Evolution of dislocation microstructure and surface slip markings. (**A**) Schematic of the proposed PSB nucleation model at the micrometer scale: (i) Within the first few loading cycles, the weakest preexisting dislocation sources on the parallel planes with the highest resolved shear stress will be activated. Those dislocations near the free surface will then escape the microcrystal creating surface steps as depicted by the straight lines. (ii) After further cyclic loading and because of the to-and-fro motion of dislocations in the crystal, dislocation tangles form on the primary slip planes. These dislocations are also entangled with secondary dislocations (not pictured). (iii) After a critical number of cycles, the dislocation tangles break free, forming primary edge dipoles. (iv) The dipoles cluster into a ladder structure and constitute a PSB nucleus because of the to-and-fro glide of screw dislocations. (v) The PSB nucleus widens and lengthens as dislocations from the surrounding matrix disentangle and become part of the PSB ladder structure. The egress of this PSB nucleus takes the form of surface markings in the direction of the primary Burgers vector. (vi) The propagation continues until the PSB occupies the entire slip region, which results in more-pronounced and -stabilized surface markings. (**B**) High-contrast SEM micrograph of the surface morphology of a Ni microcrystal cyclically loaded at a shear strain amplitude of 6.3×10^{-3} after 8 cycles (left side) and 8.5×10^6 cycles (right side). The numbered arrows indicate the observed order with which the PSB surface slip markings emerge from the surface (excluding PSBs not in the field of view during testing).

was proposed so that they could cluster and form a PSB wall. The tangles we observed also resemble those in DDD cyclic loading simulations (*24*, *37*), where these dislocation tangles were very immobile relative to dipolar loops (*37*). Thus, the limiting step in nucleating a PSB may be the number of cycles required to disentangle these dislocations and create dipolar loops. The subsequent clustering process likely happens relatively quickly after disentanglement, which is seen in DDD simulations (*24*, *28*–*30*).

We can attribute the lack of veins at the scale of our samples to the high surface-to-volume ratio (*38*). In the case of cyclic loading, the free surface makes it more difficult for the interior of the microcrystal to accumulate dislocations and accommodate plastic deformation. Because high dislocation densities are necessary to produce veins, this inhibits their formation. The dynamic stiffness curve (Fig. 1Q) provides additional evidence for this idea. This lack of veins greatly limits the number of sources available for a PSB to nucleate. This explains the large number of cycles (>10^6) necessary to nucleate a PSB and, correspondingly, the extreme fatigue lifetimes exhibited at the micrometer scale compared with those at the bulk scale.

Another potential consequence of our observations is that veins may not be a necessary condition to produce PSBs, even though they are often found in the bulk scale (*1*, *19*). A more-general requirement for the nucleation of PSBs could be the presence of enough dense dislocation tangles. Veins would meet this requirement, but the dense clustering that is characteristic of veins is arguably not a necessity. In past studies, matrix regions of fatigued fcc metals without the typical vein pattern have been observed in bulk-scale Cu and austenitic steel (*26*).

A size-dependent probabilistic model for the formation of PSBs

The physical source from which PSBs nucleate are edge dislocation dipoles that have untangled from dislocations on secondary slip systems. This can take the form of veins or dislocation tangles. Because of the large density of veins that form early during the cyclic deformation at the bulk scale, for example, there are many sources to nucleate a PSB. In small volumes, fewer sources are available to nucleate a PSB, and they take much longer to form. The yield strength of Ni microcrystals is higher than that of bulk crystals, and thus the ratio of plastic strain to overall strain is correspondingly smaller. However, in bulk single-phase fcc experiments—even at the lowest plastic strain amplitudes necessary to observe PSBs—no more than ~10^4 cycles are required to form PSBs. Therefore, the change in yield strength alone cannot account for this size effect, and a statistical explanation may be necessary.

On the basis of our PSB nucleation model, we propose a statistical approach to predict the number of cycles required for PSB formation in single crystals as a function of the crystal size. The limiting step is the formation of a local nucleus in a lamellar slip region of relatively high dislocation density. The subsequent propagation occurs shortly after the formation of a PSB nucleus. We expect a larger slip-plane area would increase the probability of PSB nucleation and reduce the number of cycles required for nucleation. We describe this using a Poisson point process distribution to represent the random nature of nucleation in a two-dimensional region. We derived the expected number of cycles (E_{cyc}) to the first nucleation event (*17*) as

$$E_{cyc} = 1 + \frac{1}{\exp\{\lambda_0[1-\exp(-D^2/D_0^2)]\} - 1} \quad (1)$$

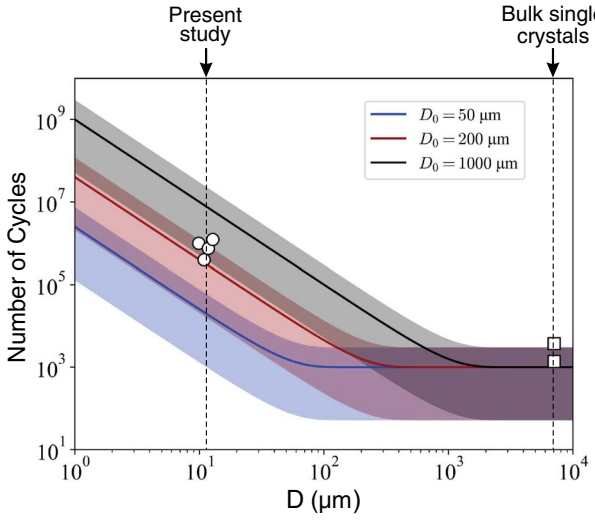


Fig. 6. Probabilistic model for PSB nucleation. Solid lines shown are the expected number of cycles for the first PSB nucleation event in a single crystal versus crystal size, D, as described by Eq. 1 with $\lambda_0 = 10^{-3}$. The shaded regions are the 90% confidence intervals. The experimentally observed numbers of cycles for the onset of saturation (which correlates with the onset of PSB formation) are also included for the microcrystals studied here and for bulk-scale crystals reported in the literature (*17*). Resolved shear strain amplitudes for data in this study range from 3.5×10^{-3} to 6.3×10^{-3}. Bulk data are reported for Ni single crystals at the onset of saturation with an orientation of $[\bar{1}49]$ (single slip) and resolved shear strain amplitudes of 3.3×10^{-3} (the upper square data point) and 4.7×10^{-3} (the lower square data point) (*35*).

where D is the edge length of the slip area, D_0 represents the critical crystal size at which a transition from bulk response to size-affected response is expected, and λ_0 is a material constant dependent on the crystallographic orientation and strain amplitude.

We plotted Eq. 1 as a function of D for $\lambda_0 = 10^{-3}$ (Fig. 6). The value of E_{cyc} corresponds to the nucleation of PSBs in bulk samples at 10^3 cycles, which is the mean value for the number of cycles to PSB nucleation at the bulk scale for the range of strain amplitudes we considered. We also considered various values for D_0. We compared saturation cycle counts for our microcrystal fatigue experiments and bulk-scale Ni single crystals (*17*, *36*). Our nucleation model captures the size effect on the PSB nucleation cycle number. The best fit for experimental results indicates that size effects start at a size $D_0 < 200$ μm. This explains the lack of observations of PSBs in many of the small-scale tests in literature that did not reach a large enough number of cycles (*15*).

Synopsis

Ni microcrystals oriented for single slip were cyclically loaded at strain amplitudes of $<1.5 \times 10^{-2}$. Using in situ surface observations, we found PSBs that nucleated locally within the microcrystal volume and then propagated gradually until they spanned the entire slip region. This gradual evolution suggests certain refinements to the conventional PSB models that idealize extrusion formation as a rapid growth process. We also found that the slip traces formed in the first ~10 cycles also identified the locations where PSBs, and thus surface cracks, will eventually form. This is important for nondestructive damage quantification (i.e., failure prediction) and for the physics-based modeling community, because simulating microstructures in fatigue is usually too computationally expensive to perform simulations past 100 cycles. A relatively large number of cycles ($>10^6$) was necessary to nucleate PSBs in microcrystals compared with bulk scale, and correspondingly, the extreme fatigue lifetimes were exhibited at the micrometer scale. On the basis of these observations, we proposed a PSB nucleation model at the micrometer scale and developed a size-dependent probabilistic model to predict the number of cycles required for PSB and subsequent crack initiation. Although the results presented here focus on pure Ni, the fundamental mechanisms identified are common to many material systems. Thus, the current insights provide an avenue to connect micrometer-scale deformation mechanisms with fatigue failures that occur at the bulk scale in metals.

Materials and methods summary

We conducted microfatigue experiments on pure Ni single crystals oriented for single slip ($[\bar{3}52]$ loading direction). We developed a high-frequency loading methodology with a microactuator. We fabricated hammerhead microcrystals using a combination of femtosecond laser milling and FIB. Microfatigue experiments were performed inside an SEM with continuously acquired micrographs during cyclic loading. We extracted and prepared post-mortem foils from the samples using FIB, and we performed bright-field TEM to characterize the dislocation microstructures. We then performed image analysis on in situ micrographs using Python-based codes. Further details can be found in the materials and methods (*17*).

REFERENCES AND NOTES

1. H. Mughbrabi, F. Ackermann, K. Herz, in *Fatigue Mechanisms*, J. Fong, Ed. (ASTM International, 1979), pp. 69–105.
2. F. Ackermann, L. P. Kubin, J. Lepinoux, H. Mughrabi, The dependence of dislocation microstructure on plastic strain ampltiude in cyclically strained copper single crystals. *Acta Metall.* **32**, 715–725 (1984). doi: 10.1016/0001-6160(84)90145-7
3. H. Kawazoe, M. Yoshida, Z. S. Basinski, M. Niewczas, Dislocation microstructures in fine-grained Cu polycrystals fatigued at low amplitude. *Scr. Mater.* **40**, 639–644 (1999). doi: 10.1016/S1359-6462(98)00484-9
4. P. Lukáš, L. Kunz, Role of persistent slip bands in fatigue. *Philos. Mag.* **84**, 317–330 (2004). doi: 10.1080/14786430310001610339
5. M. D. Sangid, The physics of fatigue crack initiation. *Int. J. Fatigue* **57**, 58–72 (2013). doi: 10.1016/j.ijfatigue.2012.10.009
6. U. Essmann, U. Gösele, H. Mughrabi, A model of extrusions and intrusions in fatigued metals I. Point-defect production and the growth of extrusions. *Philos. Mag. A* **44**, 405–426 (1981). doi: 10.1080/01418618108239541
7. J. Polák, On the Role of Point Defects in Fatigue Crack Initiation. *Mater. Sci. Eng.* **92**, 71–80 (1987). doi: 10.1016/0025-5416(87)90157-1
8. D. T. Read, Tension-tension fatigue of copper thin films. *Int. J. Fatigue* **20**, 203–209 (1998). doi: 10.1016/S0142-1123(97)00080-7
9. G. Khatibi, A. Betzwar-Kotas, V. Gröger, B. Weiss, A study of the mechanical and fatigue properties of metallic microwires. *Fatigue Fract. Eng. Mater. Struct.* **28**, 723–733 (2005). doi: 10.1111/j.1460-2695.2005.00898.x
10. T. Sumigawa, K. Byungwoon, Y. Mizuno, T. Morimura, T. Kitamura, In situ observation on formation process of nanoscale cracking during tension-compression fatigue of single crystal copper micron-scale specimen. *Acta Mater.* **153**, 270–278 (2018). doi: 10.1016/j.actamat.2018.04.061
11. C. J. Szczepanski, S. K. Jha, P. A. Shade, R. Wheeler, J. M. Larsen, Demonstration of an in situ microscale fatigue testing technique on a titanium alloy. *Int. J. Fatigue* **57**, 131–139 (2013). doi: 10.1016/j.ijfatigue.2012.08.008
12. T. Straub, M. F. Berwind, T. Kennerknecht, Y. Lapusta, C. Eberl, Small-scale multiaxial setup for damage detection into the very high cycle fatigue regime. *Exp. Mech.* **55**, 1285–1299 (2015). doi: 10.1007/s11340-015-0027-z
13. S. Lavenstein *et al.*, High frequency in situ fatigue response of Ni-base superalloy René-N5 microcrystals. *Acta Mater.* **144**, 154–163 (2018). doi: 10.1016/j.actamat.2017.10.049
14. A. Barrios, S. Gupta, G. M. Castelluccio, O. N. Pierron, Quantitative in Situ SEM High Cycle Fatigue: The Critical Role of Oxygen on Nanoscale-Void-Controlled Nucleation and Propagation of Small Cracks in Ni Microbeams. *Nano Lett.* **18**, 2595–2602 (2018). doi: 10.1021/acs.nanolett.8b00343; pmid: 29489378
15. S. Lavenstein, J. A. El-Awady, Micro-scale fatigue mechanisms in metals: Insights gained from small-scale experiments and discrete dislocation dynamics simulations. *Curr. Opin. Solid State Mater. Sci.* **23**, 100765 (2019). doi: 10.1016/j.cossms.2019.07.004
16. H. W. Höppel, P. Goik, C. Krechel, M. Göken, Ex and in situ investigations on the role of persistent slip bands and grain boundaries in fatigue crack initiation. *J. Mater. Res.* **32**, 4276–4286 (2017). doi: 10.1557/jmr.2017.313
17. Detailed materials and methods are available as supplementary materials.
18. J. Bretschneider, C. Holste, B. Tippelt, Cyclic plasticity of nickel single crystals at elevated temperatures. *Acta Mater.* **45**, 3775–3783 (1997). doi: 10.1016/S1359-6454(97)00030-X
19. A. Schwab, J. Bretschneider, C. Buque, C. Blochwitz, C. Holste, Application of electron channelling contrast to the investigation of strain localization effects in cyclically deformed fcc crystals. *Philos. Mag. Lett.* **74**, 449–454 (1996). doi: 10.1080/095008396179986

20. C. Buque, Persistent slip bands in cyclically deformed nickel polycrystals. *Int. J. Fatigue* **23**, 459–466 (2001). doi: 10.1016/S0142-1123(01)00013-5
21. H. Mughrabi, Dislocation wall and cell structures and long-range internal stress in deformed metal crystals. *Acta Metall.* **31**, 1367–1379 (1983). doi: 10.1016/0001-6160(83)90007-X
22. B.-T. Ma, C. Laird, Dislocation structures of copper single crystals for fatigue tests under variable amplitudes. *Mater. Sci. Eng. A* **102**, 247–258 (1988). doi: 10.1016/0025-5416(88)90580-0
23. T. Tabata, H. Fujita, M. A. Hiraoka, K. Onishi, Dislocation behaviour and the formation of persistent slip bands in fatigued copper single crystals observed by high-voltage electron microscopy. *Philos. Mag. A* **47**, 841–857 (1983). doi: 10.1080/01418618308243124
24. C. Erel, G. Po, T. Crosby, N. Ghoniem, Generation and interaction mechanisms of prismatic dislocation loops in FCC metals. *Comput. Mater. Sci.* **140**, 32–46 (2017). doi: 10.1016/j.commatsci.2017.07.043
25. J. Polák, T. Lepistö, P. Kettunen, Surface topography and crack initiation in emerging persistent slip bands in copper single crystals. *Mater. Sci. Eng.* **74**, 85–91 (1985). doi: 10.1016/0025-5416(85)90112-0
26. J. Polák *et al*., The role of extrusions and intrusions in fatigue crack initiation. *Eng. Fract. Mech.* **185**, 46–60 (2017). doi: 10.1016/j.engfracmech.2017.03.006
27. H. Mughrabi, Fatigue, an everlasting materials problem - Still en vogue. *Procedia Eng.* **2**, 3–26 (2010). doi: 10.1016/j.proeng.2010.03.003
28. J. Huang, N. M. Ghoniem, J. Kratochvíl, On the sweeping mechanism of dipolar dislocation loops under fatigue conditions. *Model. Simul. Mater. Sci. Eng.* **12**, 917–928 (2004). doi: 10.1088/0965-0393/12/5/010
29. T. Shimada, T. Kondo, T. Kitamura, Critical thickness for formation of fatigue dislocation structures: A discrete dislocation dynamics study. *Trans. Jpn. Soc. Mech. Eng.* **78**, 1242–1249 (2012). doi: 10.1299/kikaia.78.1242
30. C. Déprés, C. F. Robertson, M. C. Fivel, Low-strain fatigue in AISI 316L steel surface grains: A three-dimensional discrete dislocation dynamics modelling of the early cycles I. Dislocation microstructures and mechanical behaviour. *Philos. Mag.* **84**, 2257–2275 (2004). doi: 10.1080/14786430410001690051
31. J. A. El-Awady, N. Ghoniem, H. Mughrabi, "Dislocation modelling of localized plasticity in persistent slip bands" in *Proceedings of the 136th TMS Annual Meeting and Exhibition*, A. Garmestani, Ed. (The Minerals, Metals and Materials Society, 2007), pp. 23–35.
32. H. Mughrabi, Cyclic slip irreversibilities and the evolution of fatigue damage. *Metall. Mater. Trans. B* **40**, 431–453 (2009). doi: 10.1007/s11663-009-9240-4
33. Z. S. Basinski, S. J. Basinski, Fundamental aspects of low amplitude cyclic deformation in face-centred cubic crystals. *Prog. Mater. Sci.* **36**, 89–148 (1992). doi: 10.1016/0079-6425(92)90006-S
34. J. Man *et al*., Extrusion and intrusions in fatigued metals. Part 2. AFM and EBSD study of the early growth of extrusions and intrusions in 316L steel fatigued at room temperature. *Philos. Mag.* **89**, 1337–1372 (2009). doi: 10.1080/14786430902917624
35. J. Fathi Sola, R. Kelton, E. I. Meletis, H. Huang, Predicting crack initiation site in polycrystalline nickel through surface topography changes. *Int. J. Fatigue* **124**, 70–81 (2019). doi: 10.1016/j.ijfatigue.2019.02.027
36. K. Mecke, C. Blochwitz, Saturation dislocation structures in cyclically deformed nickel single crystals of different orientations. *Cryst. Res. Technol.* **17**, 743–758 (1982). doi: 10.1002/crat.2170170610
37. C. Déprés, C. F. Robertson, M. Fivel, S. Degallaix, A three dimensional discrete dislocation dynamics analysis of cyclic straining in 316L stainless steel. *Mater. Sci. Forum* **482**, 163–166 (2005). doi: 10.4028/www.scientific.net/MSF.482.163
38. A. M. Hussein, J. A. El-Awady, Quantifying dislocation microstructure evolution and cyclic hardening in fatigued face-centered cubic single crystals. *J. Mech. Phys. Solids* **91**, 126–144 (2016). doi: 10.1016/j.jmps.2016.03.012
39. D. Madisetti, S. Lavenstein, cemel-jhu/PSB-Image-Processing: Initial Release of code, version v1, Zenodo (2020); doi: 10.5281/zenodo.3978508

ACKNOWLEDGMENTS

The authors recognize C. Johnson at Drexel University for his help with TEM. The authors also appreciate discussions with H. Mughrabi of Erlangen-Nürnberg University in Germany and J. Polák of the Institute of Physics of Materials at Academy of Sciences of the Czech Republic. **Funding:** This research was sponsored by the Aerospace Materials for Extreme Environments program (program manager A. Sayir) at the U.S. Air Force Office of Scientific Research (AFOSR) under grant no. FA9550-18-1-0071 and by the National Science Foundation (NSF) CAREER award no. CMMI-1454072. D.M. also acknowledges support from the NSF Graduate Research Fellowship (GRFP) under grant no. DGE-1746891. **Authors contributions:** J.A.E.-A. conceived the idea and designed the project. S.L. and J.A.E.-A. designed the microfatigue experiments and discussed and analyzed the results. J.A.E.-A. polished and heat treated the parent Ni single crystal. S.L. fabricated the Ni microcrystals and conducted the microfatigue experiments. D.M. and S.L. developed the image analysis procedure and codes. S.L. and Y.G. developed the probabilistic model for PSB nucleation, and Y.G. performed the mathematical proofs. S.L. and J.A.E.-A. wrote the manuscript. All authors contributed to the discussion of results and provided input on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are reported in the main paper and supplementary materials. Codes for image processing of the PSB surface slip markings are available at GitHub at https://github.com/cemel-jhu/PSB-Image-Processing and Zenodo (*39*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/eabb2690/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
References (*40*–*48*)
Movie S1

12 February 2020; resubmitted 11 June 2020
Accepted 14 August 2020
10.1126/science.abb2690

RESEARCH ARTICLE SUMMARY

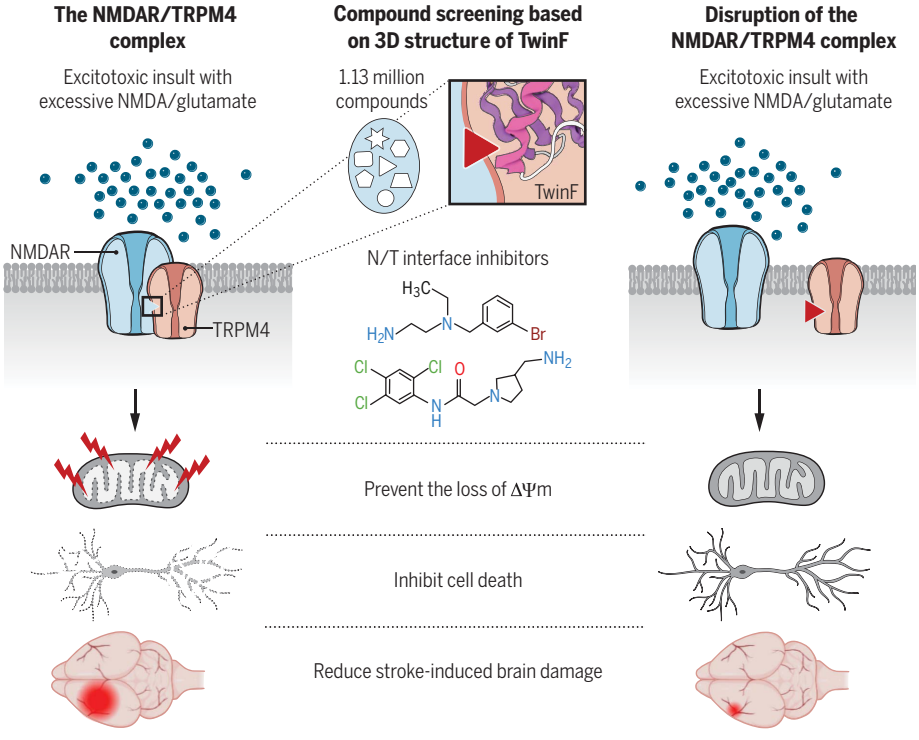
NEUROSCIENCE

Coupling of NMDA receptors and TRPM4 guides discovery of unconventional neuroprotectants

Jing Yan, C. Peter Bengtson, Bettina Buchthal, Anna M. Hagenston, Hilmar Bading*

INTRODUCTION: *N*-methyl-D-aspartate receptors (NMDARs) are glutamate-gated, calcium-permeable neurotransmitter receptors. They are fundamental to brain development, they control synaptic plasticity in the adult, and they initiate transcriptional responses needed for the consolidation of adaptive processes in the nervous system such as memory and acquired neuroprotection. However, NMDARs are also potentially harmful to neurons because they can initiate transcription shutoff pathways and can lead to mitochondrial dysfunction and even to excitotoxic cell death induced by excessive glutamate. The molecular basis of toxic NMDAR signaling is unknown, although a high intracellular calcium load has been implicated. An alternative model suggests that, depending on their location (synaptic versus extrasynaptic), NMDARs can either promote neuronal survival or cause cell death.

RATIONALE: We hypothesized that NMDARs acquire toxic features through physical interaction with one or more other proteins that may be present at extrasynaptic but not synaptic locations. Identification of NMDAR-associated proteins and the mapping of their respective interaction domains may enable the development of innovative means to disrupt a putative death signaling complex. In particular, the search for interaction interface inhibitors could yield novel compounds that—unlike classical NMDAR blockers—would bring about neuroprotection by stripping off the toxic component of extrasynaptic NMDAR signaling without compromising the physiological functions of synaptic NMDARs.



BRAIN ILLUSTRATIONS: BIORENDER.COM; GRAPHIC: N. CARY/*SCIENCE*

NMDAR/TRPM4 complex formation is required for excitotoxicity. The NMDAR/TRPM4 complex is responsible for excitotoxic damage induced by excessive glutamate. Mitochondrial dysfunction, neuronal death, and stroke-induced brain damage are schematically illustrated; ΔΨm, mitochondrial membrane potential. A structure-based compound screen using TwinF, the interface of TRPM4 in complex formation, led to the discovery of NMDAR/TRPM4 (N/T) interaction interface inhibitors, which disrupt the N/T complex and prevent excitotoxicity.

RESULTS: We found that the NMDAR subunits GluN2A and GluN2B form a complex with transient receptor potential cation channel subfamily M member 4 (TRPM4). The NMDAR/TRPM4 interaction is mediated by a 57–amino acid intracellular domain of TRPM4, termed TwinF, that is positioned just beneath the plasma membrane. TwinF interacts with I4, an evolutionarily highly conserved stretch of 18 amino acids with four regularly spaced isoleucines located within the intracellular, near-membrane portion of GluN2A and GluN2B. The NMDAR/TRPM4 complex can be disrupted by (i) expression of TwinF, which acts by competing with endogenous TRPM4 for binding to GluN2A and GluN2B, or (ii) small-molecule NMDAR/TRPM4 interaction interface inhibitors that we identified with a TwinF structure-based computational compound screen. Both TwinF and the small-molecule interface inhibitors provide robust protection against excitotoxic cell death in cultured neurons and in vivo in mouse models of neurodegeneration. They also eliminate excitotoxicity-associated transcription shutoff and mitochondrial dysfunction while leaving synaptic and extrasynaptic NMDAR-mediated currents and calcium signaling unaffected.

CONCLUSION: Our study uncovered the requirement of an NMDAR/TRPM4 complex for excitotoxicity. According to proteomics databases for synaptic proteins from mouse and human cortex and hippocampus, TRPM4 is absent from the synapse. This indicates that the NMDAR/TRPM4 complex forms extrasynaptically, which explains why extrasynaptic but not synaptic NMDARs promote death signaling. Our findings provide a conceptually new basis for therapeutic targeting of toxic NMDAR signaling, which contributes to the pathology of many neurological conditions including stroke, traumatic brain injury, Alzheimer's disease, Huntington's disease, amyotrophic lateral sclerosis, and retinal degeneration. Recombinant and small-molecule NMDAR/TRPM4 interaction interface inhibitors define a class of potent neuroprotectants with a new mode of action that renders extrasynaptic NMDARs nontoxic and eliminates their transcription shutoff signaling. Given that increased toxic signaling of extrasynaptic NMDARs is a pathomechanism shared by many neurodegenerative disorders, NMDAR/TRPM4 interaction interface inhibitors may be effective, broadly applicable therapeutics. ■

Department of Neurobiology, Interdisciplinary Center for Neurosciences (IZN), Heidelberg University, 69120 Heidelberg, Germany.
*Corresponding author. Email: bading@nbio.uni-heidelberg.de
Cite this article as J. Yan *et al.*, *Science* **370**, eaay3302 (2020). DOI: 10.1126/science.aay3302

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aay3302

RESEARCH ARTICLE

NEUROSCIENCE

Coupling of NMDA receptors and TRPM4 guides discovery of unconventional neuroprotectants

Jing Yan, C. Peter Bengtson, Bettina Buchthal, Anna M. Hagenston, Hilmar Bading*

Excitotoxicity induced by NMDA receptors (NMDARs) is thought to be intimately linked to high intracellular calcium load. Unexpectedly, NMDAR-mediated toxicity can be eliminated without affecting NMDAR-induced calcium signals. Instead, excitotoxicity requires physical coupling of NMDARs to TRPM4. This interaction is mediated by intracellular domains located in the near-membrane portions of the receptors. Structure-based computational drug screening using the interaction interface of TRPM4 in complex with NMDARs identified small molecules that spare NMDAR-induced calcium signaling but disrupt the NMDAR/TRPM4 complex. These interaction interface inhibitors strongly reduce NMDA-triggered toxicity and mitochondrial dysfunction, abolish cyclic adenosine monophosphate–responsive element–binding protein (CREB) shutoff, boost gene induction, and reduce neuronal loss in mouse models of stroke and retinal degeneration. Recombinant or small-molecule NMDAR/TRPM4 interface inhibitors may mitigate currently untreatable human neurodegenerative diseases.

N-methyl-D-aspartate (NMDA) receptors (NMDARs) are fundamental to both the physiology and pathology of the mammalian central nervous system (*1*–*4*). Calcium signals induced by synaptic NMDARs regulate neuronal plasticity and memory (*5*, *6*), initiate synapse-to-nucleus communication and gene expression changes (*7*, *8*), and lead to the buildup of a neuroprotective shield (*9*, *10*). However, NMDAR activation can also initiate transcription shutoff and cause mitochondrial dysfunction and cell death (*11*–*14*). The dichotomous roles of NMDARs explain why NMDAR antagonists, aimed at inhibiting deleterious NMDAR signaling, have proven largely unsuccessful as neuroprotectants in clinical trials, as they inherently also compromise physiological functions of NMDARs essential for cognitive abilities (*4*, *15*).

NMDARs form a complex with TRPM4

We hypothesized that the deleterious effects of NMDAR signaling are due to specific interactions between NMDARs and other proteins. We focused our search for NMDAR-interacting proteins on members of the transient receptor potential (TRP) channel family (*16*, *17*), of which the melastatin (TRPM) subgroup has previously been implicated in neurodegeneration (*18*–*20*), and identified TRPM4 as a key component of the NMDAR death signaling complex. TRPM4 is a homomeric, calcium-impermeable cationic channel activated by intracellular calcium, depolarization, and temperature (*21*). Knockdown of TRPM4 in hippocampal neurons using recombinant adeno-associated virus (rAAV)–mediated expression of TRPM4-specific short hairpin RNAs (shRNAs) significantly reduced NMDA-induced excitotoxicity but did not affect growth factor withdrawal–induced apoptosis, an NMDAR-independent form of cell death (Fig. 1, A to C). Physical coupling of the NMDAR with TRPM4 was demonstrated in coimmunoprecipitation experiments using lysates from cultured mouse hippocampal neurons and brain lysates from the mouse hippocampus and cortex (Fig. 1D). The NMDAR subunits GluN2A and GluN2B, both of which contribute to excitotoxicity (*1*, *3*, *4*, *10*, *14*, *15*, *22*, *23*), interact with TRPM4 in cultured neurons and in the mouse brain (Fig. 1D). No interaction was found between TRPM4 and the GluN1 subunit, nor between TRPM4 and the α-amino-3-hydroxy-5-methyl-4-isoxazol propionic acid (AMPA)–type glutamate receptor subunit GluA2 (Fig. 1D).

TwinF protects against excitotoxicity by disrupting the NMDAR/TRPM4 complex

We hypothesized that disruption of NMDAR/TRPM4 (N/T) interactions may eliminate NMDAR-mediated toxicity and therefore used NMDA-induced excitotoxic cell death as an assay to map the domain of TRPM4 that interacts with NMDARs. In an initial screen, a neuroprotective effect was obtained by expression of a cytoplasmic N-terminal region spanning amino acids 347 through 689 of TRPM4 (TDP347-689, where TDP denotes TRPM4-derived peptide), particularly when fused to a glycosylphosphatidylinositol (GPI) linker to create a plasma membrane anchor to mimic its localization within the native TRPM4 channel (Fig. 1, E and F, and fig. S1). Further mapping uncovered a very potent neuroprotective stretch of 57 amino acids (TDP633-689, named TwinF; see below), located just beneath the plasma membrane, which virtually eliminated NMDAR-mediated excitotoxic cell death in hippocampal neurons (Fig. 1, E and G, and fig. S1). TwinF provided robust neuroprotection in two other models of acute neurodegeneration: (i) cell death induced by oxygen-glucose deprivation (OGD) in cultured hippocampal neurons (Fig. 1H) and (ii) ischemic stroke–induced brain damage after middle cerebral artery occlusion (MCAO) in mice (Fig. 1, I to K). In the latter in vivo model, we used stereotactic injections to deliver rAAVs containing expression cassettes for TwinF or inactive controls to the mouse cortex 3 weeks prior to MCAO; brain damage was quantified 7 days after injury (Fig. 1I). The infarct volume of mice expressing TwinF in the cortex was significantly smaller than that of control mice injected with physiological saline (0.9% NaCl) or with rAAVs driving the expression of TDP536-648 or the fluorescent protein mScarlet (Fig. 1, J and K). Electrophysiological recordings from acute brain slices of mice stereotactically injected with rAAVs containing expression cassettes for tDimer and TwinF or a tDimer-only control revealed that TwinF affects neither neurons' passive properties nor their synaptic AMPA and NMDA receptor excitatory postsynaptic current (EPSC) input/output functions or kinetics (Fig. 2A and fig. S2). To probe extrasynaptic NMDAR function, we used a synaptic spillover protocol where compound EPSCs, evoked in the presence of the excitatory amino acid transporter (EAAT) blocker DL-threo-β-benzyloxyaspartate (TBOA), are prolonged by the recruitment of extrasynaptic NMDARs (*24*, *25*). Such responses were also unaffected by TwinF expression (Fig. 2B and fig. S2). The tonic NMDAR current, revealed by application of DL-2-amino-5-phosphonovaleric acid (DL-APV), was also not affected by TwinF expression (fig. S2). This tonic current is mediated by ambient glutamate that is excluded from the synapse by a protective cap of EAATs and thus selectively activates extrasynaptic NMDARs in CA1 pyramidal cells in the presence of AMPA receptor blockers (*26*, *27*). TwinF also did not affect NMDAR-mediated calcium signals in cultured hippocampal neurons and calcium signals generated by the activation of voltage-gated calcium channels (VGCCs) or by calcium release from intracellular stores (fig. S3). To investigate possible effects of TwinF on TRPM4, we used a membrane potential–sensitive fluorescence imaging plate reader (FLIPR) dye in human embryonic kidney (HEK) 293 cells transfected with an expression vector for TRPM4. Coexpression of TwinF had no effect on calcium-activated TRPM4 channel activity (fig. S4). These results indicate that the neuroprotective activity of TwinF is not the result of a blockade of NMDARs or TRPM4 channels.

Department of Neurobiology, Interdisciplinary Center for Neurosciences (IZN), Heidelberg University, 69120 Heidelberg, Germany.
*Corresponding author. Email: bading@nbio.uni-heidelberg.de

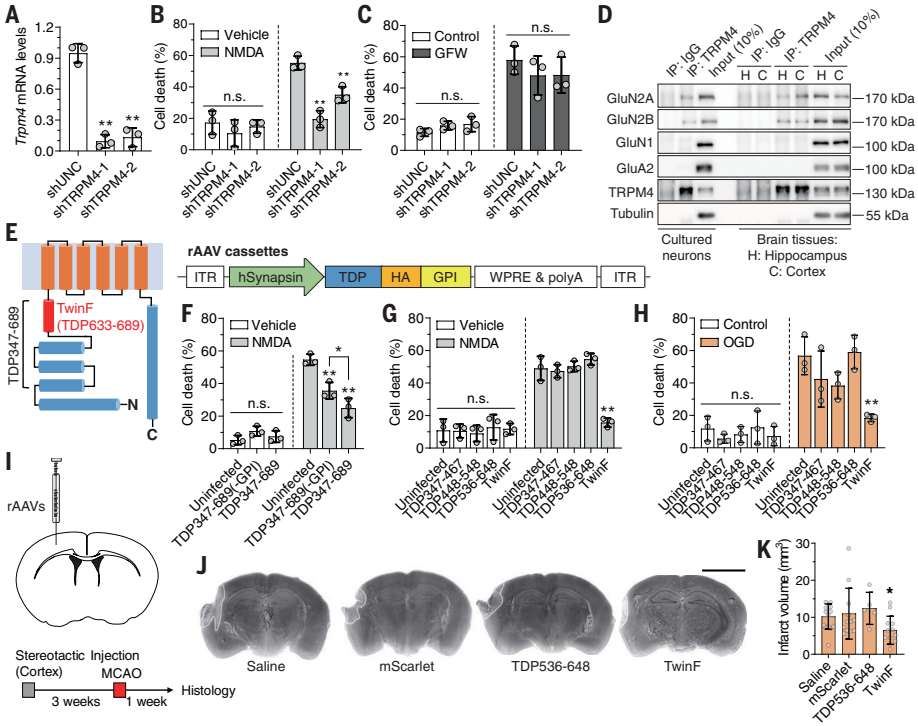


Fig. 1. TwinF protects against NMDAR-mediated toxicity. (**A** to **C**) TRPM4 knockdown in primary hippocampal neurons with rAAVs driving expression of the indicated shRNAs. (A) Knockdown efficiency assessed by quantitative reverse transcription polymerase chain reaction (qRT-PCR) from mRNA extracted on day in vitro 10 (DIV10). (B) NMDA-induced death on DIV15–16. (C) Growth factor withdrawal (GFW)–induced death in medium with (control) or without growth factors (GFW) on DIV14, with cell death assessed 3 days later. $N = 3$ independent experiments. (**D**) Coimmunoprecipitation (co-IP) of TRPM4 and NMDARs in lysates from primary neurons and brain. (**E**) Schematic of TRPM4 showing the cytosolic, near-membrane localization of TDP347-689 and TwinF (TDP633-689), and rAAV cassettes for TDP expression. (**F** to **H**) NMDA-induced death [(F) and (G)] or oxygen glucose deprivation (OGD)–induced death (H) in primary hippocampal neurons infected with rAAVs containing expression cassettes for TDPs. Cultures were challenged with NMDA (10 min) or OGD (4 hours) on DIV16, with cell death assessed 24 hours later. $N = 3$ independent experiments. (**I** to **K**) RAAV-mediated TwinF expression in the cortex protects mice from MCAO-induced brain damage. (I): Schematic illustration of rAAV injection into the cortex of mice and the timeline of MCAO treatment. (J) and (K): Images of brain slices [(J); scale bar, 3 mm] and three-dimensional analysis of infarct volume (K) after MCAO from mice injected with physiological saline or the indicated rAAVs. $N = 5$ to 15 mice in (K). Data are means ± SD; $^{*}P < 0.05$, $^{**}P < 0.01$ [one-way analysis of variance (ANOVA) followed by Holm-Sidak multiple comparisons test versus shUNC in (A); two-way ANOVA followed by Dunnett multiple-comparisons test versus shUNC or uninfected within each treatment group in (B), (C), and (F) to (H); Kruskal-Wallis ANOVA test followed by Dunnett multiple-comparisons test versus saline in (K)]; n.s., not significant.

TwinF comprises several α helices and forms a pocket-like structure with two phenylalanines, Phe^{666} and Phe^{667}, in its center (*28*–*31*) (fig. S1; see also Fig. 3, B and C). Mutation of Phe^{666} and Phe^{667} to alanine (TwinF-F_2A_2) rendered TwinF unable to provide neuroprotection against NMDAR-mediated toxicity, whereas mutation of Phe^{666} and Phe^{667} to structurally very similar tyrosines (TwinF-F_2Y_2) largely preserved TwinF's neuroprotective activity (Fig. 2C). The importance of the structural integrity of TwinF for neuroprotection became apparent in further mutational analyses. Deletion of either the N-terminal or C-terminal sequences of TwinF (i.e., TwinFΔN and TwinFΔC, respectively), as well as mutation of Cys^{654} to alanine (i.e., TwinF-CA) or Leu^{655} and Leu^{657} to proline (TwinF-L_2P_2), abolished TwinF's neuroprotective activity (fig. S1). Expression of TwinF-M5, a stretch of 57 amino acids representing the region corresponding to TwinF in TRPM5, the closest relative of TRPM4 in the TRPM family (*16*, *17*), also failed to provide neuroprotection (fig. S1). The mechanism of TwinF-mediated neuroprotection involves shielding against mitochondrial dysfunction, a hallmark of excitotoxicity and an early event en route to neuronal death (*13*, *32*–*34*). Expression of TwinF or TwinF-F_2Y_2, but not TwinF-F_2A_2, protected hippocampal neurons from NMDA-induced loss of mitochondrial membrane potential (Fig. 2, D to F) and mitochondrial reactive oxygen species production (fig. S5).

To demonstrate that disruption of the N/T interaction underlies TwinF-mediated neuroprotection, we returned to coimmunoprecipitation experiments (Fig. 2G). We cultured cortical neurons expressing hemagglutinin (HA)–tagged versions of TwinF, TwinF-F_2A_2, or TwinF-F_2Y_2; prepared whole-cell lysates; subjected them to immunoprecipitation using antibodies to either HA or TRPM4; and analyzed the immunoprecipitates for the presence of GluN2A and GluN2B. First, we detected GluN2A and GluN2B in the anti-HA immunoprecipitates from TwinF- and TwinF-F_2Y_2–expressing neurons but not in those from TwinF-F_2A_2–expressing neurons (Fig. 2G). These results indicate that TwinF and TwinF-F_2Y_2, but not the functionally inactive version TwinF-F_2A_2, interact with endogenous GluN2A and GluN2B in cortical neurons. Second, we readily detected GluN2A and GluN2B in the anti-TRPM4 immunoprecipitates from lysates of either uninfected control neurons or TwinF-F_2A_2–expressing neurons, whereas expression of TwinF or TwinF-F_2Y_2 led to a considerable reduction of GluN2A and GluN2B in the anti-TRPM4 immunoprecipitates (Fig. 2G). This confirms the observed interaction of NMDAR and TRPM4 (Fig. 1D) and indicates that (i) an inactive version of TwinF (i.e., TwinF-F_2A_2) cannot interfere with this complex, and (ii) TwinF and TwinF-F_2Y_2 can compete with endogenous TRPM4 for binding to NMDARs, thereby disrupting the endogenous N/T death complex and providing neuroprotection.

The I4 domain of GluN2A and GluN2B mediates N/T complex formation

As the binding partner of TwinF in N/T complex formation, we identified a stretch of 18 amino acids in the intracellular, near-membrane portions of GluN2A (Phe^{861} to Glu^{878}) and GluN2B (Phe^{862} to Glu^{879}) that is identical between the two subunits except for one amino acid (His^{875} in GluN2A and Ala^{876} in GluN2B) (Fig. 2H and fig. S6). This domain is highly conserved among birds and mammals (fig. S6) and was termed I4 because it contains four regularly spaced isoleucines (Ile^{863}, Ile^{867}, Ile^{871}, and Ile^{876} in GluN2A; Ile^{864}, Ile^{868}, Ile^{872}, and Ile^{877} in GluN2B) (Fig. 2H). I4 is absent from GluN1, GluN2C, GluN2D, and GluN3. Flag-tagged I4 interacted with HA-tagged TwinF in hippocampal neurons (Fig. 2I). Mutation of the two central isoleucines Ile^{868} and Ile^{872} to alanine Flag-tagged I4-I_2A_2) abolished this interaction (Fig. 2I). Attempts to coimmunoprecipitate

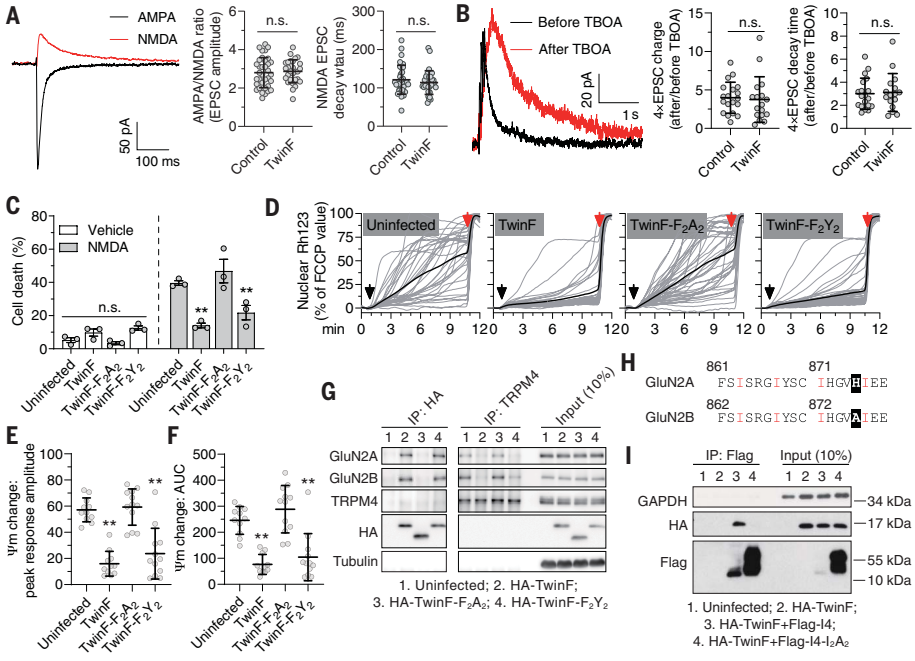


Fig. 2. TwinF prevents NMDA-induced mitochondrial dysfunction and interacts with the I4 domain of GluN2A and GluN2B. (**A**) Left: Examples of NMDAR- and AMPA receptor–mediated EPSCs recorded from a CA1 pyramidal neuron at +40 mV and −70 mV, respectively, in 1.3 mM Mg^{2+}. Right: Summary statistics for the AMPA/NMDA ratio for EPSC amplitudes and NMDAR EPSC decay weighted tau (wtau) values from bi-exponential fits. *N* = 29 to 36 cells from 14 to 17 mice. (**B**) Left: Compound NMDA EPSCs evoked by four stimuli (50-ms intervals) before and after application of TBOA. Right: Summary statistics for the normalized charge transfer and decay times (10 to 90%) of the responses. *N* = 17 to 19 cells from 6 or 7 mice. (**C**) NMDA-induced cell death in primary hippocampal neurons challenged with NMDA on DIV16. (**D** to **F**) NMDA-induced mitochondrial membrane potential (Ψm) changes in hippocampal neurons. (D): Rh123 fluorescence changes in response to a 10-min NMDA application (black arrow) normalized to their maximal response to carbonylcyanide-*p*-(trifluoromethoxy)-phenylhydrazone (FCCP; red arrow) from a single coverslip for each condition. Gray, individual cells; black, their mean. (E) and (F): Summary statistics for the peak response (E) and area under the curve (AUC) during NMDA application (F). *N* = 11 or 12 coverslips from three independent experiments. (**G**) Co-IP of HA-tagged TwinF and its mutants or of endogenous TRPM4 in primary cortical neurons with GluN2A and GluN2B, respectively. (**H**) Amino acid sequence of the I4 domain within mouse GluN2A and GluN2B. The four regularly spaced isoleucines are shown in red. Abbreviations: A, Ala; C, Cys; E, Glu; G, Gly; F, Phe; H, His; I, Ile; R, Arg; S, Ser; V, Val; Y, Tyr. One mismatch (H versus A) is highlighted. (**I**) Co-IP of HA-TwinF and GluN2B-derived, Flag-tagged I4 or its mutant, I4-I_2A_2. Data are means ± SD; $^*P < 0.05$, $^{**}P < 0.01$ [independent-samples two-tailed *t* tests in (A) and (B); two-way ANOVA followed by Dunnett multiple-comparisons test versus uninfected within each treatment group in (C); one-way ANOVA followed by Dunnett multiple-comparisons test versus uninfected in (E) and (F)].

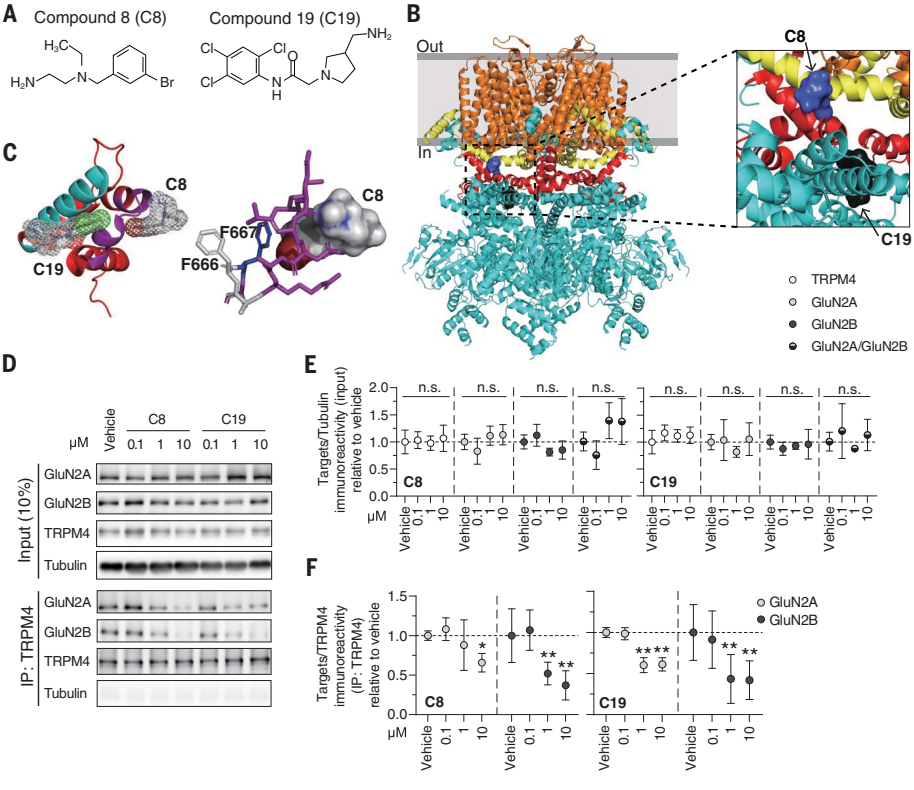


Fig. 3. Identification of novel compounds that disrupt the NMDAR/TRPM4 interaction. (**A**) Two-dimensional chemical structures of compound 8 (C8) and compound 19 (C19). (**B**) Surface representations of C8 (blue) and C19 (black) at their predicted interaction sites with the TwinF (Asn^{633} to Pro^{689}; red) region of TRPM4 (PDB ID 6BCO) shown as a PyMOL cartoon helical structure with transmembrane (orange), TRP (yellow), and cytoplasmic (cyan) domains (see also movies S1 and S2). (**C**) Left: Docking of C8 and C19 (PyMOL licorice/mesh representation) to the Ala^{658}-Ala^{668} region (purple) of TwinF (red) and also of C19 to an adjacent N-terminal helix (Asp^{614} to His^{632}; cyan). Right: Binding of C8 (PyMOL surface representation) to a basket-like structure (PyMOL licorice representation) formed by Ala^{658}-Arg^{664} (purple) and Phe^{667} (blue). Phe^{666} and Phe^{667} represent the two neighboring, functionally relevant phenylalanines of TwinF. (**D** to **F**) Co-IP of the NMDAR/TRPM4 complex with anti-TRPM4 antibody in primary cortical neurons pre-incubated with C8 or C19. (D): Input and TRPM4 immunoprecipitates probed with the indicated antibodies. (E) and (F): Immunoreactivity of TRPM4, GluN2A, GluN2B, and GluN2A/GluN2B normalized to tubulin for input samples (E) and of GluN2A and GluN2B normalized to TRPM4 for IP samples (F) relative to vehicle (dashed lines). Data are means ± SD; *N* = 3 independent experiments. $^*P < 0.05$, $^{**}P < 0.01$ (two-way ANOVA followed by Dunnett multiple-comparisons test versus vehicle).

endogenous TRPM4 with I4 generated variable results as I4 expression compromised the health of neurons (fig. S7). This form of toxicity may be the result of I4 mimicking—through binding to the TwinF domain of endogenous TRPM4—the interaction of NMDAR and TRPM4 in excitotoxicity. Coexpression of TwinF to scavenge I4 (that otherwise could bind to TwinF of endogenous TRPM4) or knockdown of TRPM4 in hippocampal neurons prevented I4-induced cell death (fig. S7).

TwinF structure-based compound screen identifies neuroprotective N/T interaction interface inhibitors

The potent neuroprotective activity of TwinF both in vitro and in vivo raises hopes for possible future clinical applications. However, a therapy regime involving stereotactic delivery of rAAVs to the brain appears unsuitable for humans. A preferable treatment would involve small molecules administered orally or intravenously. We reasoned that it may be possible to disrupt the N/T death complex with small molecules that bind to TwinF and block the N/T interaction interface. We carried out a TwinF structure-based computational drug screen, which revealed 192 candidate molecules. Compound 8 and compound 19 were selected on the basis of internal strain calculations and docking scores, and they are presented here as prototypes of a new class of neuroprotective small molecules. They have simple structures with no known resemblance to any other type of drug, and thus also chemically define a new class of bioactive substances (Fig. 3, A to C, movies S1 and S2, and models S1 and S2).

To establish that compound 8 and compound 19 are functional interaction interface inhibitors, we first used coimmunoprecipitation to assess their ability to disrupt the complex of endogenous GluN2A or GluN2B with TRPM4 in hippocampal neurons. Compound 8 and compound 19 reduced, in a dose-dependent manner, the interactions of GluN2A and GluN2B with TRPM4 (Fig. 3, D and F) while leaving the expression levels of TRPM4, GluN2A, and GluN2B unaffected (Fig. 3, D and E). Both compounds are potent neuroprotectants. They reduced, in a dose-dependent manner, NMDA-induced cell death in hippocampal neurons [compound 8, half-maximal inhibitory concentration (IC_{50}) 2.1 μM; compound 19, IC_{50} 1.1 μM; Fig. 4A] but did not affect NMDAR-independent apoptotic cell death triggered by growth factor withdrawal (fig. S8). Measurements of the mitochondrial membrane potential revealed that both compound 8 and compound 19 (10 μM) virtually eliminated NMDA-induced mitochondrial dysfunction in hippocampal neurons (Fig. 4, B to D). Even high concentrations (100 μM) of compound 8 were well tolerated by the neurons and produced no signs of cell damage (Fig. 4A). However, compound 19 applied at high concentrations compromised the health of the neurons [half-maximal effective concentration (EC_{50}) 34.7 μM; Fig. 4A] and was therefore used for in vitro experiments only.

We investigated possible effects of compound 8 and compound 19 on NMDA-induced calcium signaling, which could explain their neuroprotective activities. In contrast to the NMDAR antagonist DL-APV, which completely blocked NMDA-induced calcium signals, neither compound 8 nor compound 19 (10 μM) affected NMDA-induced calcium signals in hippocampal neurons in the absence (Fig. 5, A to D) or in the presence of VGCC blockers (fig. S9), nor did they affect calcium signals in hippocampal neurons generated by the activation of VGCCs or by calcium release from intracellular stores (fig. S9). Synaptic activity–induced calcium signals triggered by the γ-aminobutyric acid type A ($GABA_A$) receptor blocker gabazine were also unaffected by the compounds (Fig. 5, E to H). $GABA_A$ receptor blockers remove inhibition from the hippocampal network, leading to periodic bursts of action potentials (APs) (*13*) and calcium entry through both NMDARs and VGCCs (*7*–*9*, *35*). Using HEK293 cells that express recombinant NMDARs, we found no effect of compound 8 and compound 19 on NMDAR currents (fig. S10) and no effect on NMDAR activation kinetics (fig. S11). To directly assess the impact of compound 8 and compound 19 on NMDAR currents and NMDAR-mediated calcium influx in neurons independent of contributions from depolarization-activated calcium entry via VGCCs, we used single-cell patch-clamp analysis in the presence of VGCC blockers. We found no effect of either compound on NMDAR-mediated inward currents in neurons clamped at −70 mV or on the NMDAR conductance at depolarized potentials (fig. S12). Because compound 8 was used for animal experiments (Fig. 6), we additionally performed single-cell patch-clamp analysis combined with bis-Fura2 calcium imaging in cultured hippocampal neurons as well as patch-clamp recordings of synaptic and extrasynaptic NMDAR currents in acute hippocampal slices and found little effect of compound 8 (figs. S12 and S13). At −70 mV in low magnesium, but not at +40 mV in physiological magnesium concentrations, compound 8 slightly increased synaptic NMDAR rise time quantified from exponential fits (although not decay time) and caused a small but not significant decrease in the holding current (fig. S13). This may reflect a minor blockade or loss of extrasynaptic NMDARs in low-magnesium recording conditions. These results indicate on the whole that, analogously to TwinF (see Fig. 2, A and B, and figs. S2 and S3), neither compound 8 nor compound 19 are blockers of NMDAR currents or NMDAR-induced calcium signals in hippocampal neurons in primary cultures or acute slices or in HEK293 cell expression systems. However, we cannot rule out the possibility that features of NMDARs or other recording paradigms not assessed in this study are affected by the new compounds or by TwinF.

To investigate possible effects of compound 8 and compound 19 on TRPM4 channels, we again used a FLIPR membrane potential–sensitive dye in HEK293 cells transfected with an expression vector for TRPM4. Compound 8 had no effect on calcium-activated TRPM4 channel activity (fig. S4). Using compound 19,

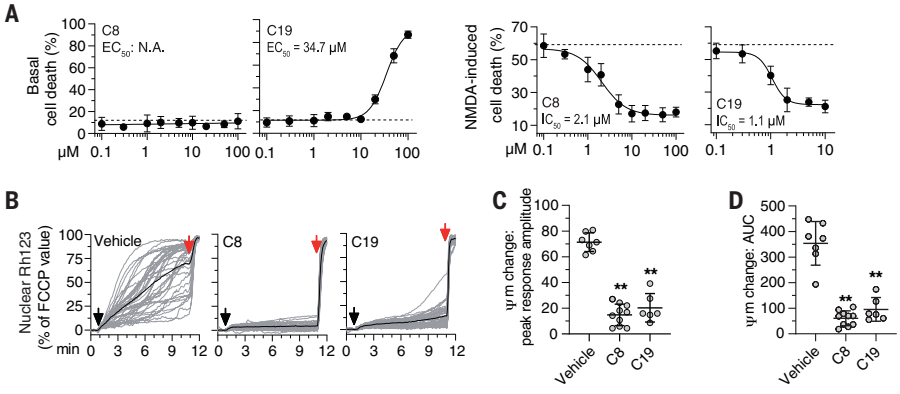


Fig. 4. Compound 8 and compound 19 prevent NMDA-induced cell death and mitochondrial dysfunction. (**A**) Left: Determination of the EC_{50} of C8 and C19 for basal cell death (24 hours of exposure to compounds) relative to vehicle (dashed lines). Right: Determination of the IC_{50} for NMDA-induced cell death in primary hippocampal neurons relative to vehicle. N.A., not applicable. Data are means ± SD; N = 3 independent experiments in each case. (**B** to **D**) Effects of C8 and C19 on NMDA-induced Ψm changes in primary hippocampal neurons. (B): Rh123 fluorescence changes in response to NMDA application (black arrow) normalized to the maximal response to FCCP (red arrow) from a single coverslip for each condition. Gray, individual cells; black, their mean. (C): Summary statistics for the peak response. (D): Summary statistics for AUC during NMDA application. Data are means ± SD; N = 4 to 10 coverslips from three independent experiments. IC_{50} estimates were generated from logistic fits of the Hill equation to dose-response relationships. $^{**}P < 0.01$ (one-way ANOVA followed by Dunnett multiple-comparisons test versus vehicle).

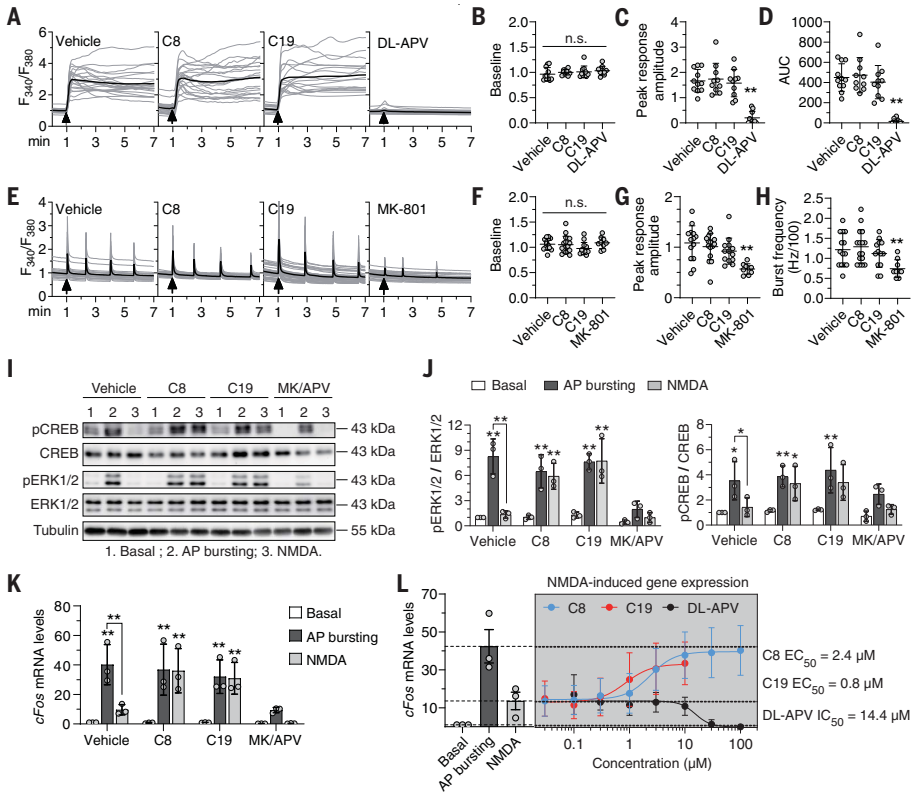


Fig. 5. Compound 8 and compound 19 boost downstream signaling responses to NMDA without affecting calcium signals. (**A** to **H**) Effects of C8, C19, and DL-APV or MK-801 on NMDA-induced calcium influx during a 6-min NMDA application [(A) to (D)] or on calcium transients arising from gabazine-induced action potential (AP) bursting [(E) to (H)] in primary hippocampal neurons. (A) and (E): Fura2-AM imaging traces of NMDA-induced cytosolic calcium signals from a single coverslip for each condition. Gray, individual cells; black, their mean. (B) to (D): Quantitative analysis of baseline (B), peak response amplitude (C), and AUC (D) for NMDA-induced calcium influx. (F) to (H): Quantitative analysis of baseline (F), peak response amplitude (G), and burst frequency (H) for gabazine-triggered AP bursting. In (B) to (D) and (F) to (H), N = 9 to 16 coverslips from three or four independent experiments. (**I** to **L**) The effects of C8, C19, or MK-801 and DL-APV combined (MK/APV) on ERK1/2 or CREB phosphorylation and *cFos* expression in response to either bicuculline-induced AP bursting or bath application of NMDA in primary hippocampal neurons. (I) and (J): Immunoblots and quantitative analysis of pERK1/2 and pCREB for a 15-min stimulation. (K) and (L): qRT-PCR analysis of *cFos* mRNA expression levels for a 1-hour stimulation. In (L), *cFos* expression in vehicle-treated cultures from (K) is shown again for reference for the EC_{50} analysis of C8 or C19 enhancement (left) and the IC_{50} analysis of DL-APV inhibition of NMDA-induced *cFos* expression (right). In (J) to (L), N = 3 independent experiments. Data are means ± SD. *P < 0.05, **P < 0.01 [one-way ANOVA followed by Dunnett multiple-comparisons test versus vehicle in (B) to (D) and (F) to (H); two-way ANOVA followed by Dunnett multiple-comparisons test versus basal within each treatment group in (J) and (K)]. In (L), IC_{50} estimates were generated from logistic fits of the Hill equation to dose-response relationships.

however, we did observe—although only at high concentrations—an inhibition of the TRPM4 channel; the IC_{50} obtained was 12.9 μM (fig. S4), which is more than one order of magnitude higher than the IC_{50} of compound 19 for inhibition of NMDA-induced cell death (IC_{50} 1.1 μM; Fig. 4A). This effect of compound 19 on TRPM4 may be explained by the position of its contact site within the central pore-lining structure of the TRPM4 channel (fig. S14 and movies S3 and S4). In contrast, the TwinF binding pocket for compound 8 is located on the surface of TRPM4 (fig. S14 and movies S3 and S4).

We next investigated the possibility that compound 8 and compound 19 mitigate the transcriptional deregulation that is associated with NMDAR-mediated toxicity and is primarily caused by shutoff of the transcription factor CREB (cyclic adenosine monophosphate–responsive element–binding protein) and poor activation of ERK1/2 signaling (*13*, *36*). Treatment of hippocampal neurons with compound 8 or compound 19 converted otherwise toxic NMDA bath application into a potent inducer of CREB activity, as assessed in immunoblot analyses using antibodies that recognize the Ser^{133}-phosphorylated (i.e., activated) form of CREB (Fig. 5, I and J, and fig. S15). Similarly, compound 8 and compound 19 amplified NMDA-induced ERK1/2 activation, measured using activation state–specific antibodies, although the effect was detected at 15 min but not at 60 min after stimulation (Fig. 5, I and J, and fig. S15). The boost of NMDA-induced CREB and ERK1/2 activation by compound 8 and compound 19 translates into a robust enhancement of NMDA-induced gene expression. Compound 8 and compound 19 dose-dependently transformed the NMDA bath application–induced expression of the immediate early genes (IEGs) *cFos*, *Bdnf*, *Npas4*, *Arc*, and *Atf3* to levels that were comparable to those achieved by AP bursting (Fig. 5, K and L, and fig. S15), which requires synaptic NMDAR activation (*2*, *7*–*9*, *13*). The NMDAR blockers MK-801 and DL-APV reduced AP bursting and the accompanying calcium signals (*35*) (Fig. 5E) as well as the subsequent activation of CREB, ERK1/2, and IEG expression (Fig. 5, I to K, and fig. S15), whereas neither compound 8 nor compound 19 had an effect (Fig. 5, I to K, and fig. S15). Thus, compound 8 and compound 19 detoxify NMDAR signaling: They eliminate the CREB shutoff pathway and restore ERK1/2 activation and IEG induction while sparing the synaptic activity–driven, transcription-promoting activities of NMDARs.

N/T interaction interface inhibitor provides neuroprotection in vivo

We finally investigated whether small-molecule N/T interaction interface inhibitors could provide neuroprotection in vivo. We focused on compound 8 because compound 19 at high concentrations can compromise the health of neurons (Fig. 4A). We also observed that compound 8 can protect cultured neurons even after prolonged (20 to 24 hours) exposure to NMDA (fig. S16), making this compound particularly suitable for therapeutic purposes in animals. We used two mouse models of neurodegeneration: ischemic stroke induced by MCAO, and retinal ganglion cell (RGC) degeneration induced by the intravitreal injection of NMDA. To establish a suitable treatment regime, we used the disruption of the N/T complex, assessed by coimmunoprecipitations from brain lysates (as in Fig. 1D), as an indicator of compound delivery to the target brain area (Fig. 6, A to C). The expression levels of TRPM4, GluN2A, and GluN2B remained unchanged during the course of the treatment with compound 8 (Fig. 6, A and B). However, relative to controls, we found a reduction in GluN2A/TRPM4 complex formation of 26% at 2 hours and of 36% at 6 hours, and a reduction in GluN2B/TRPM4 complex formation of 38% at 2 hours and of 36% at 6 hours, after a single intraperitoneal (i.p.) injection of compound 8 at 40 mg per kg of body weight (Fig. 6, A and C). Twenty-four hours after i.p. injection of compound 8, the

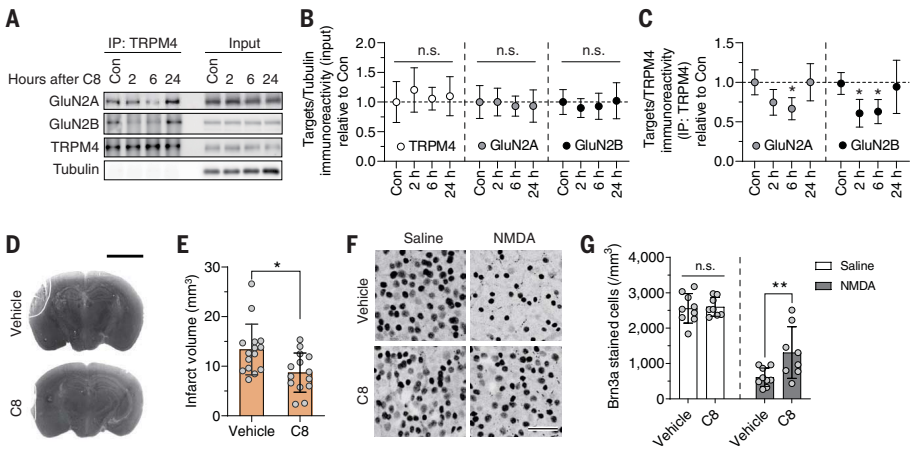


Fig. 6. Compound 8 protects mice from MCAO-induced brain damage and NMDA-induced retinal ganglion cell loss. (**A** to **C**) Co-IP of the NMDAR/TRPM4 complex with anti-TRPM4 antibody from cortical lysates obtained from control mice or from mice that had received an intraperitoneal injection of C8 2 hours, 6 hours, or 24 hours earlier. (A): Images of immunoblots. (B) and (C): Immunoreactivity of TRPM4, GluN2A, and GluN2B to tubulin for input samples (B) and of GluN2A and GluN2B to TRPM4 immunoreactivity for IP samples relative to control (Con) (C); $N = 6$ independent experiments. (**D** and **E**) C8 reduces MCAO-induced brain damage. (D): Silver-stained coronal sections 1 week after MCAO in mice that received vehicle or C8. Scale bar, 3.0 mm. (E): Three-dimensional analysis of infarct volume in whole ischemic brains; $N = 14$ or 15 mice. (**F** and **G**) C8 reduces retinal ganglion cell (RGC) degeneration as assessed 1 week after intravitreal injection of mice with NMDA. (F): Images from whole-mount retinas. Scale bar, 50 μm. (G): Quantitative analysis of immunolabeling for Brn3a to mark live RGCs; $N = 8$ or 9 mice. Data are means ± SD. $^*P < 0.05$, $^{**}P < 0.01$ [one-way ANOVA followed by Dunnett multiple-comparisons test versus control in (B) and (C); independent-samples two-tailed t test in (E); two-way ANOVA followed by Sidak multiple-comparisons test versus vehicle within each treatment group in (G)].

GluN2A/TRPM4 and GluN2B/TRPM4 complexes had reformed (Fig. 6, A and C). We used a multiple injection regime for compound 8 for our in vivo neuroprotection experiments (see supplementary materials for details). We observed a significant reduction in brain damage in the MCAO stroke model (Fig. 6, D and E) and robust protection against NMDA-induced degeneration of RGCs (Fig. 6, F and G).

Our findings reveal that excitotoxicity is due to a physical coupling of NMDARs to TRPM4, in contrast to the current view of "NMDAR overactivation and calcium overload" (*1*, *3*, *37*). TRPM4 is absent from the synapse (synaptome databases: SynProt, SynSysNet, SynaptomeDB) (*38*), which indicates that the N/T complex forms extrasynaptically. This resolves the long-standing enigma that, depending on their location (synaptic versus extrasynaptic), NMDARs either promote neuronal survival or cause cell death (*13*). Disruption of the N/T complex with TwinF, compound 8, or compound 19 strips off the toxic component of extrasynaptic NMDARs, thereby rendering them functionally similar to synaptic NMDARs, which are inert to treatment with N/T interaction interface inhibitors and remain unaltered in their physiological signaling (Fig. 7). This offers an opportunity for novel neuroprotective therapies. It now appears possible to specifically target toxic NMDAR signaling with recombinant proteins or small molecules that act as N/T interaction interface inhibitors while sparing the critical physiological functions of NMDARs. This pharmacological mode of action represents a radically new therapeutic principle that may prove far superior to the use of classical NMDAR blockers that compromise cognitive functions and, except for low-dose memantine (*39*), have failed clinically (*4*, *15*). TwinF and compound 8 hold great potential as effective therapeutics for currently untreatable human neurodegenerative diseases.

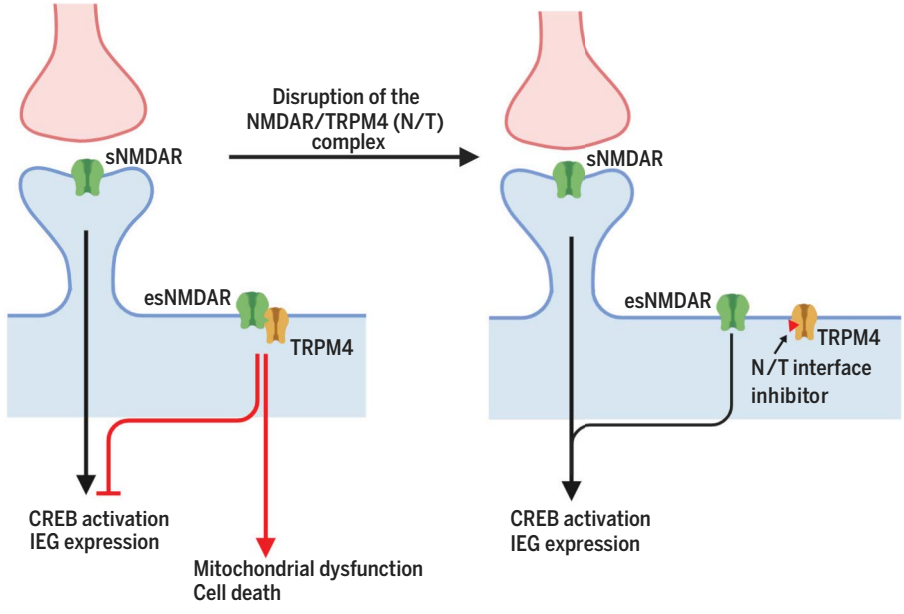


Fig. 7. Model for neuroprotection by NMDAR/TRPM4 interaction interface inhibitors. Schematic illustration of the disruption of the extrasynaptically localized NMDAR/TRPM4 (N/T) complex with N/T interface inhibitors, which renders extrasynaptic NMDARs (esNMDARs) nontoxic. N/T interface inhibitors also eliminate the CREB shutoff pathway, thereby converting the transcription-inhibiting signaling of esNMDARs to a mode resembling that of synaptic NMDARs (sNMDARs) that strongly promote immediate early gene (IEG) expression.

REFERENCES AND NOTES

1. P. Paoletti, C. Bellone, Q. Zhou, NMDA receptor subunit diversity: Impact on receptor properties, synaptic plasticity and disease. *Nat. Rev. Neurosci.* **14**, 383–400 (2013). doi: 10.1038/nrn3504; pmid: 23686171
2. G. E. Hardingham, H. Bading, The Yin and Yang of NMDA receptor signalling. *Trends Neurosci.* **26**, 81–89 (2003). doi: 10.1016/S0166-2236(02)00040-1; pmid: 12536131
3. A. Lau, M. Tymianski, Glutamate receptors, neurotoxicity and neurodegeneration. *Pflugers Arch.* **460**, 525–542 (2010). doi: 10.1007/s00424-010-0809-1; pmid: 20229265
4. K. K. Ogden, S. F. Traynelis, New advances in NMDA receptor pharmacology. *Trends Pharmacol. Sci.* **32**, 726–733 (2011). doi: 10.1016/j.tips.2011.08.003; pmid: 21996280
5. R. G. M. Morris, NMDA receptors and memory encoding. *Neuropharmacology* **74**, 32–40 (2013). doi: 10.1016/j.neuropharm.2013.04.014; pmid: 23628345
6. A. Volianskis *et al.*, Long-term potentiation and the role of N-methyl-D-aspartate receptors. *Brain Res.* **1621**, 5–16 (2015). doi: 10.1016/j.brainres.2015.01.016; pmid: 25619552

7. A. M. Hagenston, H. Bading, Calcium signaling in synapse-to-nucleus communication. *Cold Spring Harb. Perspect. Biol.* **3**, a004564 (2011). doi: 10.1101/cshperspect.a004564; pmid: 21791697
8. A. E. West, M. E. Greenberg, Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, a005744 (2011). doi: 10.1101/cshperspect.a005744; pmid: 21555405
9. H. Bading, Nuclear calcium signalling in the regulation of brain function. *Nat. Rev. Neurosci.* **14**, 593–608 (2013). doi: 10.1038/nrn3531; pmid: 23942469
10. G. E. Hardingham, H. Bading, Synaptic versus extrasynaptic NMDA receptor signalling: Implications for neurodegenerative disorders. *Nat. Rev. Neurosci.* **11**, 682–696 (2010). doi: 10.1038/nrn2911; pmid: 20842175
11. D. W. Choi, Ionic dependence of glutamate neurotoxicity. *J. Neurosci.* **7**, 369–379 (1987). doi: 10.1523/JNEUROSCI.07-02-00369.1987; pmid: 2880938
12. S. A. Lipton, P. A. Rosenberg, Excitatory amino acids as a final common pathway for neurologic disorders. *N. Engl. J. Med.* **330**, 613–622 (1994). doi: 10.1056/NEJM199403033300907; pmid: 7905600
13. G. E. Hardingham, Y. Fukunaga, H. Bading, Extrasynaptic NMDARs oppose synaptic NMDARs by triggering CREB shut-off and cell death pathways. *Nat. Neurosci.* **5**, 405–414 (2002). doi: 10.1038/nn835; pmid: 11953750
14. M. P. Parsons, L. A. Raymond, Extrasynaptic NMDA receptor involvement in central nervous system disorders. *Neuron* **82**, 279–293 (2014). doi: 10.1016/j.neuron.2014.03.030; pmid: 24742457
15. H. Bading, Therapeutic targeting of the pathological triad of extrasynaptic NMDA receptor signaling in neurodegenerations. *J. Exp. Med.* **214**, 569–578 (2017). doi: 10.1084/jem.20161673; pmid: 28209726
16. C. Montell, L. Birnbaumer, V. Flockerzi, The TRP channels, a remarkably functional family. *Cell* **108**, 595–598 (2002). doi: 10.1016/S0092-8674(02)00670-0; pmid: 11893331
17. M. M. Moran, H. Xu, D. E. Clapham, TRP ion channels in the nervous system. *Curr. Opin. Neurobiol.* **14**, 362–369 (2004). doi: 10.1016/j.conb.2004.05.003; pmid: 15194117
18. M. Aarts *et al.*, A key role for TRPM7 channels in anoxic neuronal death. *Cell* **115**, 863–877 (2003). doi: 10.1016/S0092-8674(03)01017-1; pmid: 14697204
19. H.-S. Sun *et al.*, Suppression of hippocampal TRPM7 protein prevents delayed neuronal death in brain ischemia. *Nat. Neurosci.* **12**, 1300–1307 (2009). doi: 10.1038/nn.2395; pmid: 19734892
20. B. Schattling *et al.*, TRPM4 cation channel mediates axonal and neuronal degeneration in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nat. Med.* **18**, 1805–1811 (2012). doi: 10.1038/nm.3015; pmid: 23160238
21. P. Launay *et al.*, TRPM4 is a Ca^{2+}-activated nonselective cation channel mediating cell membrane depolarization. *Cell* **109**, 397–407 (2002). doi: 10.1016/S0092-8674(02)00719-5; pmid: 12015988
22. E. Morikawa *et al.*, Attenuation of focal ischemic brain injury in mice deficient in the ε1 (NR2A) subunit of NMDA receptor. *J. Neurosci.* **18**, 9727–9732 (1998). doi: 10.1523/JNEUROSCI.18-23-09727.1998; pmid: 9822733
23. J. von Engelhardt *et al.*, Excitotoxicity in vitro by NR2A- and NR2B-containing NMDA receptors. *Neuropharmacology* **53**, 10–17 (2007). doi: 10.1016/j.neuropharm.2007.04.015; pmid: 17570444
24. B. A. Clark, S. G. Cull-Candy, Activity-dependent recruitment of extrasynaptic NMDA receptor activation at an AMPA receptor-only synapse. *J. Neurosci.* **22**, 4428–4436 (2002). doi: 10.1523/JNEUROSCI.22-11-04428.2002; pmid: 12040050
25. N. A. Lozovaya, M. V. Kopanitsa, Y. A. Boychuk, O. A. Krishtal, Enhancement of glutamate release uncovers spillover-mediated transmission by N-methyl-D-aspartate receptors in the rat hippocampus. *Neuroscience* **91**, 1321–1330 (1999). doi: 10.1016/S0306-4522(98)00638-1; pmid: 10391439
26. N. Lozovaya *et al.*, Protective cap over CA1 synapses: Extrasynaptic glutamate does not reach the postsynaptic density. *Brain Res.* **1011**, 195–205 (2004). doi: 10.1016/j.brainres.2004.03.023; pmid: 15157806
27. K. Le Meur, M. Galante, M. C. Angulo, E. Audinat, Tonic activation of NMDA receptors by ambient glutamate of non-synaptic origin in the rat hippocampus. *J. Physiol.* **580**, 373–383 (2007). doi: 10.1113/jphysiol.2006.123570; pmid: 17185337
28. J. Guo *et al.*, Structures of the calcium-activated, non-selective cation channel TRPM4. *Nature* **552**, 205–209 (2017). doi: 10.1038/nature24997; pmid: 29211714
29. P. A. Winkler, Y. Huang, W. Sun, J. Du, W. Lü, Electron cryo-microscopy structure of a human TRPM4 channel. *Nature* **552**, 200–204 (2017). doi: 10.1038/nature24674; pmid: 29211723
30. H. E. Autzen *et al.*, Structure of the human TRPM4 ion channel in a lipid nanodisc. *Science* **359**, 228–232 (2018). doi: 10.1126/science.aar4510; pmid: 29217581
31. J. Duan *et al.*, Structure of full-length human TRPM4. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2377–2382 (2018). doi: 10.1073/pnas.1722038115; pmid: 29463718
32. R. J. White, I. J. Reynolds, Mitochondrial depolarization in glutamate-stimulated neurons: An early signal specific to excitotoxin exposure. *J. Neurosci.* **16**, 5688–5697 (1996). doi: 10.1523/JNEUROSCI.16-18-05688.1996; pmid: 8795624
33. A. F. Schinder, E. C. Olson, N. C. Spitzer, M. Montal, Mitochondrial dysfunction is a primary event in glutamate neurotoxicity. *J. Neurosci.* **16**, 6125–6133 (1996). doi: 10.1523/JNEUROSCI.16-19-06125.1996; pmid: 8815895
34. O. Vergun, J. Keelan, B. I. Khodorov, M. R. Duchen, Glutamate-induced mitochondrial depolarisation and perturbation of calcium homeostasis in cultured rat hippocampal neurones. *J. Physiol.* **519**, 451–466 (1999). doi: 10.1111/j.1469-7793.1999.0451m.x; pmid: 10457062
35. C. P. Bengtson, M. Kaiser, J. Obermayer, H. Bading, Calcium responses to synaptically activated bursts of action potentials and their synapse-independent replay in cultured networks of hippocampal neurons. *Biochim. Biophys. Acta* **1833**, 1672–1679 (2013). doi: 10.1016/j.bbamcr.2013.01.022; pmid: 23360982
36. A. Ivanov *et al.*, Opposing role of synaptic and extrasynaptic NMDA receptors in regulation of the extracellular signal-regulated kinases (ERK) activity in cultured rat hippocampal neurons. *J. Physiol.* **572**, 789–798 (2006). doi: 10.1113/jphysiol.2006.105510; pmid: 16513670
37. K. Szydlowska, M. Tymianski, Calcium, ischemia and excitotoxicity. *Cell Calcium* **47**, 122–129 (2010). doi: 10.1016/j.ceca.2010.01.003; pmid: 20167368
38. A. Bayés *et al.*, Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLOS ONE* **7**, e46683 (2012). doi: 10.1371/journal.pone.0046683; pmid: 23071613
39. S. A. Lipton, Paradigm shift in neuroprotection by NMDA receptor blockade: Memantine and beyond. *Nat. Rev. Drug Discov.* **5**, 160–170 (2006). doi: 10.1038/nrd1958; pmid: 16424917

ACKNOWLEDGMENTS

We thank I. Bünzli-Ehret for help with the preparation of hippocampal cultures; U. Weiss for technical assistance; M. Freichel (Department of Pharmacology, Heidelberg University) for providing a TRPM4 plasmid; M. Hackbusch (Department of Medical Biometry, Institute of Medical Biometry and Informatics, Heidelberg University) for help with the biostatistical and biometrical planning of the study; A. Marklew and colleagues at Charles River Chesterford Research Park for their help with HEK293 cell electrophysiology; and A. Zalewski (Proteros GmbH) for technical assistance. The schematic in Fig. 7 and parts of the graphical abstract were created with BioRender.com. All images and movies illustrating TRPM4 were generated using PDB ID 6BCO. **Funding:** Supported by the Deutsche Forschungsgemeinschaft (DFG) Deutsch-Israelische Projektkooperation (BA 1007/7-1), DFG/Agence Nationale de la Recherche (ANR) Project GLAD (BA 1007/11-1), European Research Council (ERC) Advanced Grant 233024, DFG Sonderforschungsbereich (SFB) 1134 and SFB 1158, and DFG Forschergruppe FOR 2289. H.B. is a member of the Excellence Cluster CellNetworks at Heidelberg University. **Author contributions:** H.B. initiated the project, led the study, selected compound 8 and compound 19 from the list of candidates, analyzed the data, and wrote the paper; J.Y. carried out all experiments except for the MCAO and electrophysiological experiments and analyzed the data; B.B. carried out the MCAO experiments and analyzed the data; J.Y. and B.B. carried out the RGC degeneration experiments; A.M.H. and J.Y. carried out imaging experiments and analyzed the data; C.P.B. carried out electrophysiological experiments and analyzed the data. **Competing interests:** H.B. and J.Y. are named inventors of the patent application for a novel class of neuroprotectants (PCT/EP2018/078577); H.B. is cofounder of FundaMental Pharma GmbH, Heidelberg, Germany, which licenses compound 8 and compound 19 under the names Brophenexin and Aliudanexin, respectively; H.B. and J.Y. are shareholders of FundaMental Pharma GmbH. **Data and materials availability:** All data are available in the main text or the supplementary materials. All raw statistical data are provided in the statistics summary. All plasmids used in this study will be freely available for biomedical research under a material transfer agreement. Compound 8 and compound 19 are available from FundaMental Pharma GmbH, Heidelberg. The PDB files illustrating predictions of compound 8 and compound 19 binding to TRPM4 (PDB ID 6BCO) are included as model S1 and model S2.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/eaay3302/suppl/DC1
Materials and Methods
Statistics Summary
Figs. S1 to S16
References (*40*–*49*)
Movies S1 to S4
Models S1 and S2

10 June 2019; resubmitted 5 May 2020
Accepted 13 August 2020
10.1126/science.aay3302

RESEARCH ARTICLE

BATTERIES

Black phosphorus composites with engineered interfaces for high-rate high-capacity lithium storage

Hongchang Jin[1]*, Sen Xin[2,3]*, Chenghao Chuang[4], Wangda Li[3], Haiyun Wang[1], Jian Zhu[5], Huanyu Xie[1], Taiming Zhang[1], Yangyang Wan[1], Zhikai Qi[1], Wensheng Yan[6], Ying-Rui Lu[7], Ting-Shan Chan[7], Xiaojun Wu[1], John B. Goodenough[3], Hengxing Ji[1]†, Xiangfeng Duan[8]†

High-rate lithium (Li) ion batteries that can be charged in minutes and store enough energy for a 350-mile driving range are highly desired for all-electric vehicles. A high charging rate usually leads to sacrifices in capacity and cycling stability. We report use of black phosphorus (BP) as the active anode for high-rate, high-capacity Li storage. The formation of covalent bonds with graphitic carbon restrains edge reconstruction in layered BP particles to ensure open edges for fast Li^+ entry; the coating of the covalently bonded BP-graphite particles with electrolyte-swollen polyaniline yields a stable solid–electrolyte interphase and inhibits the continuous growth of poorly conducting Li fluorides and carbonates to ensure efficient Li^+ transport. The resultant composite anode demonstrates an excellent combination of capacity, rate, and cycling endurance.

Lithium ion batteries (LIBs) are increasingly important for diverse applications, including electrical vehicles. However, today's batteries can only provide a limited power density (e.g., ~100 to 300 W kg^{-1} at the cell level) and typically require a relatively long charging time (hours or longer) for safe operation (*1*, *2*). To improve charging rate, specific energy, and battery lifetime, anode materials with a high Li storage capacity, high rate capability, and high electrochemical stability are essential. Binary composites consisting of a Li^+ insertion host (e.g., graphite) for fast and stable Li storage and an alloying element (e.g., silicon) for high lithiation capacity have been developed (*3*) with a reversible capacity of 517 mA·hour g^{-1} (with respect to the composite mass) and an areal capacity of >3.3 mA·hour cm^{-2}, showing commercial viability for next-generation LIBs (*4*). However, such a capacity is only achieved at a relatively low current density of 0.26 A g^{-1} with a full charging time of ~2 hours. To be competitive with 5-min refuel time for conventional combustion engine vehicles, all-electric road vehicles require LIB cells that reach a full charge of 350 W·hour kg^{-1} after a comparable duration. Achieving these goals requires anode materials that can be charged to a specific capacity of 350 to 400 mA·hour g^{-1} at a charge–discharge current density of >5 A g^{-1} (*5*). To this end, it is essential to develop an electrode material simultaneously featuring high theoretical capacity along with the excellent electron conductivity and Li^+ diffusivity that are necessary for rapid charge (*6*).

Layered black phosphorus (BP) exhibits several attractive features for high-rate, high-capacity Li storage. Through a three-electron alloying reaction with Li^+, BP can theoretically deliver a gravimetric capacity of 2596 mA·hour g^{-1} (*7*, *8*), which is only bettered by Si (4200 mA·hour g^{-1}) and Li metal (3860 mA·hour g^{-1}) (*9*). The large capacity of BP helps offset its relatively high voltage loss versus (Li/Li^+) (~0.7 V on average) to render a high specific energy density according to the equation $E = V(q) \times Q(I_{dis})$, where $V(q)$ is the mean cell voltage versus state of charge q and $Q(I_{dis})$ is the capacity density for a given discharge current, with I_{dis} calculated as dq/dt. Additionally, the electrical conductivity of BP is ~300 S m^{-1} (*10*), four orders of magnitude greater than that of silicon (6.7×10^{-2} S m^{-1}) (*11*); the Li^+ diffusion barrier along the zigzag direction of layered BP is only 0.08 eV (*12*),

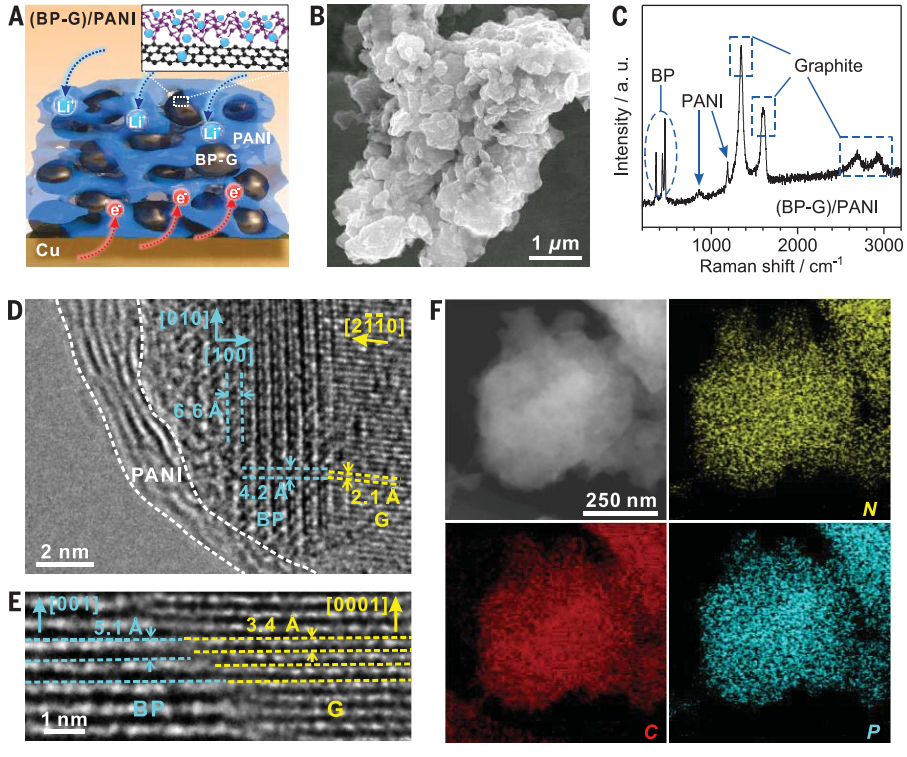


Fig. 1. Structure of (BP-G)/PANI. (**A**) Schematic of (BP-G)/PANI. (**B** and **C**) SEM image (B) and Raman spectrum (C) of (BP-G)/PANI. a.u., arbitrary units. (**D**) TEM image showing the crystalline domains of BP and a graphite flake covered with PANI. (**E**) High-resolution TEM image showing the merge of basal planes of BP and graphite. Every two BP layers match with three graphene layers. (**F**) Dark-field TEM image and P, C, and N elemental maps of (BP-G)/PANI.

[1]Hefei National Laboratory for Physical Sciences at the Microscale, CAS Key Laboratory of Materials for Energy Conversion, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China. [2]CAS Key Laboratory of Molecular Nanostructure and Nanotechnology, Beijing National Laboratory for Molecular Sciences (BNLMS), Institute of Chemistry, Chinese Academy of Sciences (CAS), Beijing 100190, China. [3]Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712, USA. [4]Department of Physics, Tamkang University, Tamsui 251, New Taipei City, Taiwan. [5]State Key Laboratory for Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China. [6]National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230029, China. [7]National Synchrotron Radiation Research Center, 300 Hsinchu, Taiwan. [8]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jihengx@ustc.edu.cn (He.J.); xduan@chem.ucla.edu (X.D.)

considerably lower than that in silicon (0.58 eV) (*13*). These combined features have invited investigations of the rate of Li^+ diffusion in BP. Studies to date show a faster Li^+ diffusion in bulk BP than in silicon or other conventional anode materials. However, an edge atom reconstruction near the zigzag diffusion channel of BP nanoflakes hinders the kinetics of Li^+ transfer across the surface (*14*). Additionally, the volume change of BP during charge–discharge cycles renders the solid–electrolyte interphase (SEI) unstable, leading to poor cycling performance (*7*).

We present a BP-graphite (BP-G) hybrid with a covalently bonded BP-G interface, to prevent edge reconstruction and ensure efficient Li^+ insertion and diffusion, and a thin polyaniline (PANI) polymer gel coating swollen by electrolytes to prevent the continued formation and buildup of less-conductive Li fluorides and carbonates, leading to a stable SEI that is more conductive for Li^+. The designed (BP-G)/PANI composite with optimized interface for Li^+ conduction delivers a combination of high rate and high capacity with robust cycling stability.

Structural characterization of (BP-G)/PANI composite

The (BP-G)/PANI composites (Fig. 1A) were fabricated by ball milling a mixture of BP and graphite and then in situ polymerization of PANI (materials and methods, figs. S1 and S2, and table S1). The scanning electron microscopy (SEM) image of the composite shows aggregates of (BP-G)/PANI particles (Fig. 1B). The Raman spectrum shows intense bands of the A_g^1, B_{2g}, and A_g^2 vibration modes of BP at 363, 440, and 467 cm^{-1}, respectively (Fig. 1C) (*15*). The Raman bands centered at 850 and 1200 cm^{-1} are attributed to PANI (*16*), and those at 1330 and 1590 cm^{-1} are the D and G bands of graphite, respectively (*17*).

The transmission electron microscopy (TEM) image (Fig. 1D) obtained at the edge of a (BP-G)/PANI particle presents three domains with different crystalline features. The crystal domain in the particle center shows lattice fringes with d-spacings of 6.6 and 4.2 Å, corresponding to the (100) and (010) lattice planes of BP crystal, respectively (*8*). The crystal domain on the right-hand side shows fringes with a d-spacing of 2.1 Å, corresponding to the $(10\bar{1}0)$ plane of graphite (*18*). PANI is amorphous and covers the surface of the BP-G domain (Fig. 1D).

The lattice-resolved TEM image also reveals that the d-spacings across the basal planes of BP and graphite are 5.1 and 3.4 Å, respectively, with every two BP layers matching with three graphene layers (Fig. 1E). Dislocations are evident at the interface due to mismatched layer numbers. Nonetheless, with weak van der Waals interaction across the basal planes, the strain caused by the dislocations may be released near the grain boundary, and the crystal structures away from the grain boundary largely retain nearly ideal structures (*19*). These results indicate

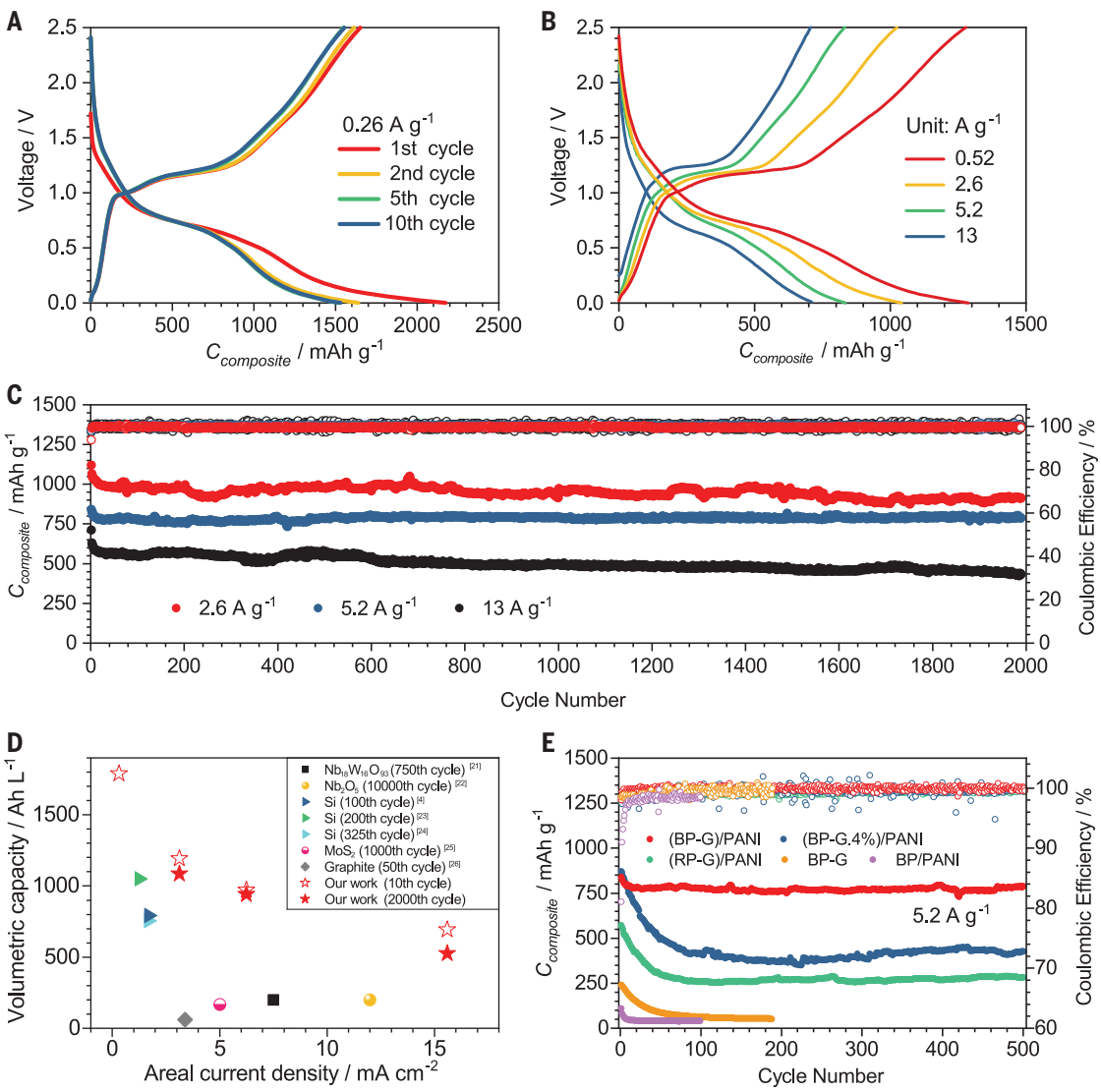


Fig. 2. Electrochemical performance of (BP-G)/PANI. (**A** and **B**) Galvanostatic discharge and charge profiles of different cycles (A) and measured at different current densities (B). (**C**) Cycling performance of (BP-G)/PANI measured at current densities of 2.6, 5.2, and 13 A g^{-1}. (**D**) Volumetric performance metrics of (BP-G)/PANI electrode compared with various state-of-the-art anodes. The volumetric capacities are calculated based on the gravimetric capacity and packing densities of the electrodes, including carbon black and binder. The data of the 10th cycle, which were considered the performance of an electrode after its initial aging process, were obtained from (A) and (C). The data of the 2000th cycle, which were considered the performance of an electrode at the end of its service life in practical LIBs, were obtained from (C). (**E**) Cycling performance of (BP-G)/PANI composite compared with different control samples.

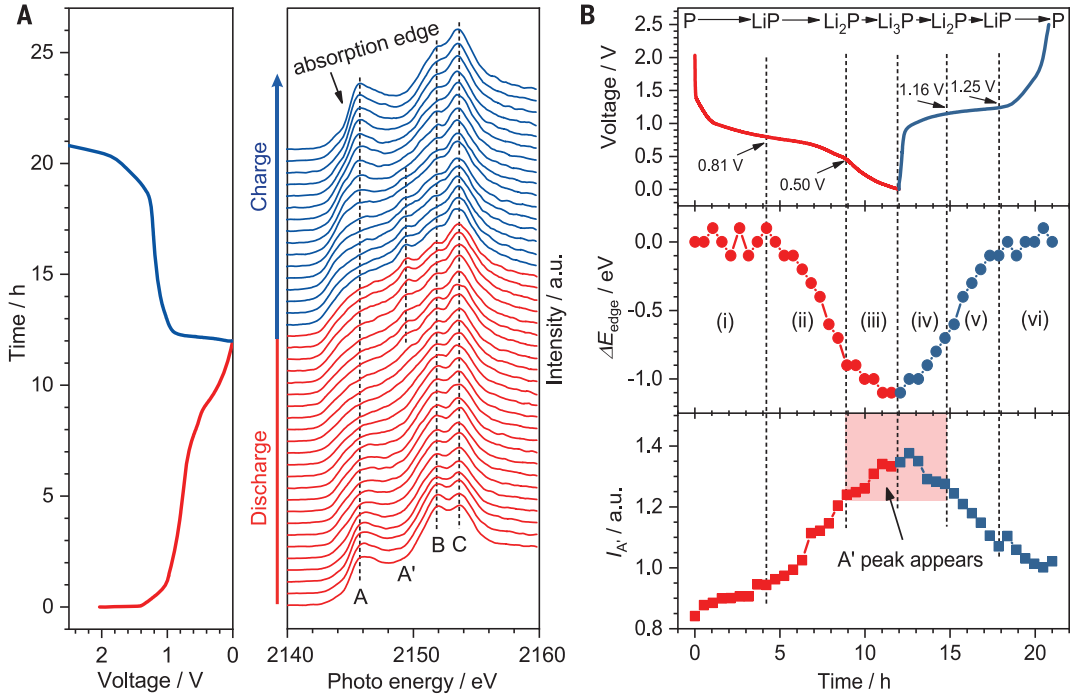


Fig. 3. In situ XAS tracking of the structural evolution of BP-G electrodes. (**A**) A series of XAS near-edge spectra for the P *K*-edge as a function of discharge–charge cycle time (*t*), recorded from a BP-G electrode during the first galvanostatic lithiation–delithiation cycle at 0.13 A g^{-1}. The state of charge corresponds to the time scale of the voltage profile (left). (**B**) Relative shift of the absorption edge (ΔE_{edge}) and absorption intensity at peak A′ as a function of discharge–charge time. The dashed lines divide six regions (indicated by lowercase roman numerals) based on the slope of the $\Delta E_{edge} - t$ curve. The red area marks the region where peak A′ appears.

that the basal planes of BP and graphite are intimately merged together.

The dark-field TEM image (Fig. 1F) shows the (BP-G)/PANI is composed of primarily particles with sizes of ~500 nm, and the elemental maps show a rather consistent distribution of P, C, and N with a morphology almost identical to that in the dark-field TEM image, indicating a uniform PANI coating.

Li storage performance of (BP-G)/PANI

The synthesized (BP-G)/PANI (containing 65 wt % BP, 16 wt % graphite, and 19 wt % PANI) (materials and methods) was characterized by the constant current method in a voltage range of 0.001 to 2.5 V versus (Li/Li^+). The gravimetric capacities and current densities were quantified by normalizing to the total mass of the entire composite. The (BP-G)/PANI delivered initial discharge and charge capacities of 2170 and 1650 mA·hour g^{-1} at 0.26 A g^{-1} (Fig. 2A), with an initial Coulombic efficiency of 76.0%. At the 10th cycle, the (BP-G)/PANI showed a reversible discharge capacity of 1520 mA·hour g^{-1}. The discharge profiles measured during different cycles show a plateau at ~0.75 V (Fig. 2A), which is characteristic of the formation of Li_xP (*20*). This plateau remains resolvable at increased discharge current densities of 0.52, 2.6, 5.2, and 13 A g^{-1} (Fig. 2B), indicating excellent reaction kinetics between BP and Li^+. To probe the dominant charge storage kinetics in (BP-G)/PANI, we analyzed the capacity contribution from the diffusion-controlled process and the capacitive process (fig. S3 and supplementary text S1). The pseudocapacitive process contributes a substantial (nearly constant) fraction of the overall capacity at low or high rates, which is beneficial to high-power operation.

We tested the cycling performance at different charge–discharge current densities (Fig. 2C). The (BP-G)/PANI anode (at an areal loading of 1.2 mg cm^{-2}) showed a reversible capacity of 910 mA·hour g^{-1} after cycling at 2.6 A g^{-1} for 2000 cycles, a reversible capacity of 790 mA·hour g^{-1} after cycling at 5.2 A g^{-1} for 2000 cycles, and a reversible capacity of 440 mA·hour g^{-1} after cycling at 13 A g^{-1} for 2000 cycles. The ability to charge an anode material at 13 A g^{-1} with a reversible capacity of 440 mA·hour g^{-1} promises advanced LIBs that require <10-min charging to a cell-level energy density of ~350 W·hour kg^{-1} (*5*). When the areal loadings were increased to 2.2 and 3.6 mg cm^{-2}, the (BP-G)/PANI retained gravimetric capacities of 980 and 750 mA·hour g^{-1}, respectively (fig. S4).

The electrochemical performance was obtained with a conventional slurry electrode structure consisting of (BP-G)/PANI, carbon black, and binder (materials and methods) at an overall packing density of 1.49 ± 0.08 g cm^{-3}, which is close to that of a commercial graphite anode (table S2). The porosity of the (BP-G)/PANI anode is 0.29 ± 0.04 (supplementary text S2). Once experimental gravimetric capacity and electrode packing density are considered, the (BP-G)/PANI presents a volumetric capacity of 1090 A·hour $liter^{-1}$ at 3.1 mA cm^{-2} after 2000 cycles and 530 A·hour $liter^{-1}$ at 15.6 mA cm^{-2} after 2000 cycles (Fig. 2D). Such high-rate performance, especially after extended cycling, exceeds those of traditional carbon anodes or advanced silicon anodes and is comparable with those of known high-rate materials (Fig. 2D and table S3) (*4*, *21*–*26*), yet with a considerably higher capacity.

Taking the volumetric capacities of the whole anode (*21*), we can project the energy density with respect to the volume of the (BP-G)/PANI anode, to reach up to 1940 W·hour $liter^{-1}$ at a power density of 50 kW $liter^{-1}$ (fig. S5 and supplementary text S3). A prototype full cell comprised of a $LiCoO_2$ cathode and a (BP-G)/PANI anode shows comparable performance to that in half-cells (fig. S6).

To elucidate the roles of BP, graphite, and PANI in the (BP-G)/PANI for Li storage, we prepared four different composites (materials and methods): (i) red phosphorus (RP)–based ternary composites, (RP-G)/PANI; (ii) (BP-G.4%)/PANI (containing less graphite content, 4 wt %); (iii) BP-G; and (iv) BP/PANI. Their structures and cyclic voltammetry (CV) profiles are shown in figs. S7 and S8, and their cycling performances are compared in Fig. 2E. The discharge capacity of (BP-G)/PANI is around three times that of (RP-G)/PANI at the same rate, suggesting that BP contributes to at least 70% of the capacity of (BP-G)/PANI. The reversible capacities of the composites decreased with decreasing graphite content in the order (BP-G)/PANI > (BP-G.4%)/PANI > BP/PANI (Fig. 2E), even though (BP-G)/PANI contained the lowest mass of BP (65 wt %) and thus the lowest theoretical capacity, which

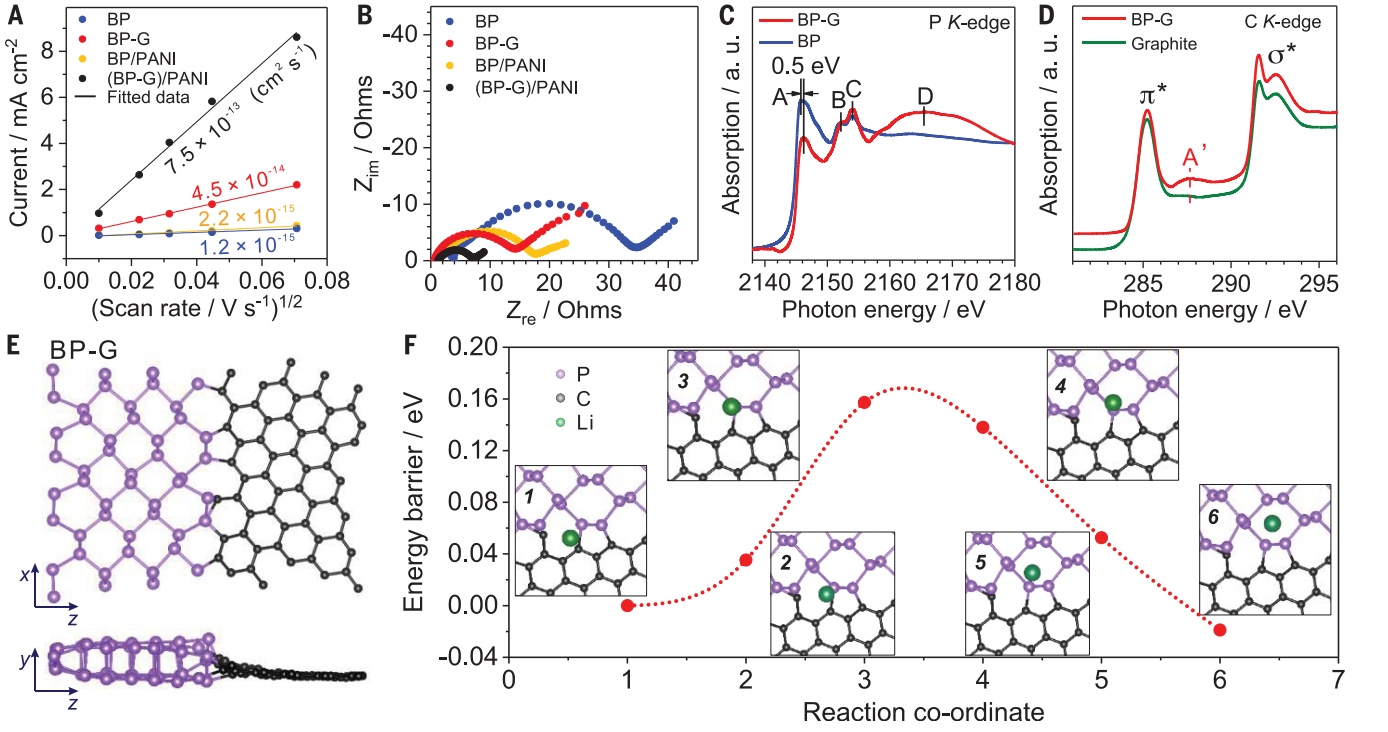


Fig. 4. Charge transfer properties and structure of BP-G. (**A** and **B**) D_{Li} values derived from the CV profiles (A) and Nyquist plots (B) of (BP-G)/PANI, BP-G, BP/PANI, and BP. (**C** and **D**) P *K*-edge XAS spectra of BP-G and BP (C) and C *K*-edge XAS spectra of BP-G and graphite (D). (**E**) Schematic of BP-G hybrid structure with in plane P–C bonds. (**F**) The minimum energy pathways for Li^+ to diffuse through the BP-G boundary. The results shown in (E) and (F) were obtained from the DFT calculations.

demonstrates the important role of graphite in facilitating Li storage in BP. Moreover, the reversible capacity of (BP-G)/PANI is ~10 times that of BP-G, demonstrating the role of PANI in preserving the cycling stability of BP.

The role of BP-G bonding

To further probe the role of graphite in Li storage, we performed in situ x-ray absorption spectroscopy (XAS) measurements on BP-G electrodes (Fig. 3). The voltage profile shows discharge and charge plateaus at ~0.75 and ~1.20 V, respectively (Fig. 3A), which is much higher than the (de)lithiation voltage of graphite [~0.1 V versus (Li/Li^+)]. Therefore, Li reacts predominantly with BP during in situ XAS measurements. At the beginning of discharge, the absorption edge is located at 2144.3 eV, with three edge peaks, A, B, and C, evident in the P *K*-edge XAS spectrum at 2146.2, 2152.0, and 2153.7 eV, respectively (Fig. 3A). In the whole discharge–charge process, the positions of peaks B and C do not change, whereas the absorption edge shifts [defined by $\Delta E_{edge} = E(t) - E(0)$, where $E(t)$ and $E(0)$ are the positions of absorption edges in the spectra of intermediate and initial states, respectively], and a new peak, A′, appears at the discharge–charge time (t) of 8.9 to 14.8 hours.

The $\Delta E_{edge} - t$ curve is plotted together with the voltage profile of the BP-G electrode in Fig. 3B. According to the slope of the curve, there are three regions in the discharge process: (region i) $t < 4.2$ hours, where ΔE_{edge} is ~0.0 eV; (region ii) $4.2 < t < 8.9$ hours, where ΔE_{edge} evolves rapidly to −0.9 eV; and (region iii) $8.9 < t < 11.9$ hours, where ΔE_{edge} evolves slowly to −1.1 eV at the end of discharge (t = 11.9 hours). The voltages at t of 4.2, 8.9, and 11.9 hours are 0.81, 0.50, and 0.01 V, respectively, matching those of LiP, Li_2P, and Li_3P in the discharge profile (*20*) and the XRD measurements (fig. S9). The red shift of the absorption edge of the P *K*-edge spectrum indicates electron acceptance of P atoms (*27*). The little change in ΔE_{edge} in region i indicates that the phase change between P and LiP involves only a weak bonding between Li and P atoms (*28*) and high reversibility of the phase change between P and LiP (*7*). The formation of Li_2P is accompanied by a large evolution of ΔE_{edge} (Fig. 3B, region ii) because of the substantial structural distortion of BP and the increased number of Li–P bonds in Li_2P (*20*). Li_3P has a hexagonal layered structure (P63/mmc) consisting of alternatively arranged Li and Li_2P layers (*29*); the change of ΔE_{edge} is only 0.2 eV from Li_2P to Li_3P, but the phase change from Li_2P to Li_3P leads to the appearance and growth of an A′ peak at 2149.5 eV.

During charging, the ΔE_{edge} shows an essentially symmetric profile with a slow increase to −0.7 eV at 14.8 hours (Fig. 3B, region iv), then a more rapid increase to around 0.0 eV at 17.9 hours (Fig. 3B, region v), and then stays at 0.0 eV from 17.9 to 21.0 hours (Fig. 3B, region vi). The intensity of peak A′ shows a similar mirror evolution as a function of discharge and charge time. The reversible increase and decrease of ΔE_{edge} and the intensity of peak A′ with a symmetric profile during the charge and discharge process indicate that the phase change between BP and Li_3P in BP-G is highly reversible. The change of the XAS spectra with the state of charge of the bare BP was much smaller (fig. S10), suggesting a considerably lower lithiation level of BP than BP-G. Together, these in situ studies demonstrate that the presence of graphite in the BP-G hybrid allows a deeper and more reversible (de)lithiation of BP, which contributes to the improved Li storage performance.

We conducted CV and electrical impedance spectroscopy (EIS) studies and compared the results with those for several control electrodes, including (BP-G)/PANI, BP-G, BP/PANI, and BP (fig. S8). The derived Li^+ diffusion coefficients (D_{Li}) in BP-G (4.5×10^{-14} cm^2 s^{-1}) and (BP-G)/PANI (7.5×10^{-13} cm^2 s^{-1}) are one to two orders of magnitude higher than that in BP (1.2×10^{-15} cm^2 s^{-1}) and BP/PANI (2.2×10^{-15} cm^2 s^{-1}) (Fig. 4A and supplementary text S4). The charge transfer resistance (R_{ct}) in BP-G (13.7 ohms) and (BP-G)/PANI (5.9 ohms) is considerably lower than that in BP (31.1 ohms) (Fig. 4B, fig. S8, and table S4). The higher D_{Li}

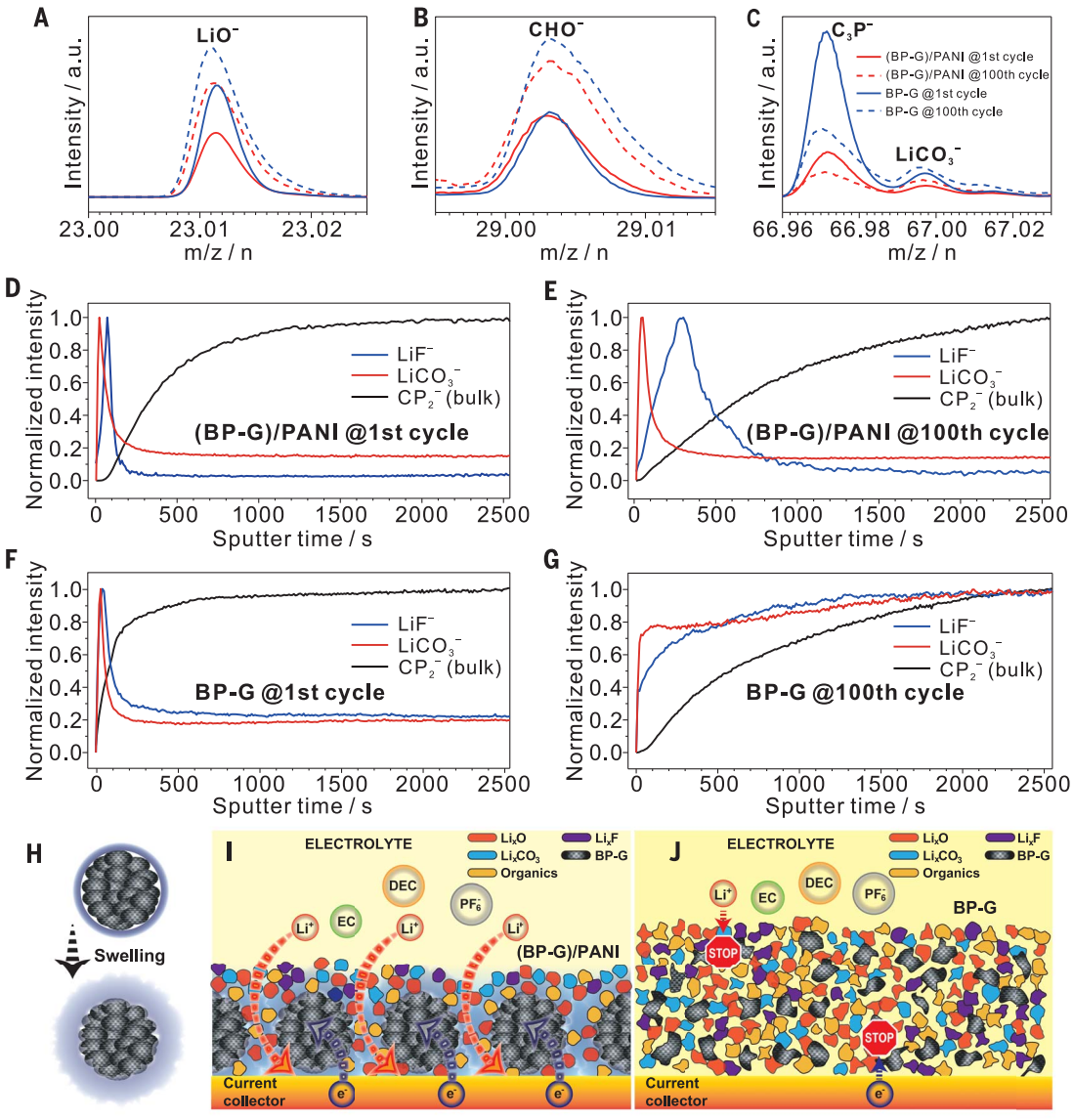


Fig. 5. Interfacial study of (BP-G)/PANI and BP-G anodes. (**A** to **C**) ToF-SIMS spectra of secondary ion fragments LiO^- (A) CHO^- (B), and C_3P^- and $LiCO_3^-$ (C) collected from the cycled (BP-G)/PANI and BP-G electrodes. (**D** to **G**) Normalized depth profiles of some representative secondary ion fragments obtained from the (BP-G)/PANI anode after the first cycle (D) and after the 100th cycle (E), and from the BP-G anode after the first cycle (F) and after the 100th cycle (G). (**H**) Schematic of electrolyte-swollen BP-G coated with PANI. (**I** and **J**) Schematics of the (BP-G)/PANI electrode, in which the PANI coating helps retain a stable SEI (I) to prevent continuous formation of poorly conducting Li fluorides and carbonates deep into the BP-G particles (J).

and lower R_{ct} of BP-G hybrids are consistent with the in situ XAS results showing that graphite improves the completeness and reversibility of the lithiation reaction, which contributes to the improved rate capability and cycling stability. Beyond the electrochemical benefits, the efficient ion transport and reduced overall impedance of the composite together with the high thermal conductivity of the BP and G may also help minimize heat generation and accelerate heat dissipation during high-rate operation (supplementary text S5), which is beneficial for cycling stability and battery safety.

Previous theoretical calculations have indicated that the wide channel and low diffusion barrier along the zigzag direction of BP are responsible for the high D_{Li} in BP (*12*). However, theoretical calculations (*30*) also suggest that reconstruction at the BP edges (the curving of the basal planes, as also confirmed by TEM observation) (*14*) results in a narrower gap for Li^+ diffusion and thus limits its performance. We believe that the formation of P–C bonds between BP and graphite flakes (during the ball milling process) mitigates the edge reconstruction problem, and this is supported by the TEM images (Fig. 1, D and E) and XAS analysis (Fig. 4, C and D) and also the density functional theory (DFT) calculation (Fig. 4, E and F). The P *K*-edge XAS spectra show four distinguishable peaks: A, B, C, and D (Fig. 4C). The peak A of BP-G (2146.2 eV) has a photon energy higher than that of BP (2145.7 eV) (*31*), which is in agreement with the P *L*-edge spectra (fig. S11A and supplementary text S6) and is more likely attributable to charge transfer from a P atom to a C or H atom rather than P–O bonding (with a typical photon energy of >2147.0 eV) (*32*). Additionally, elemental analysis (table S1) shows that both BP and BP-G have negligible amounts of hydrogen. Hence, peak A is exclusively ascribed to P–C bonding. A broad peak (D) at ~2165.0 eV appears in the XAS of BP-G (Fig. 4C) but not in that of BP; this has been attributed to the distortion of the P site by the neighboring C atoms in P-doped diamond (*33*). Similarly, the peak A′ in the C *K*-edge XAS (Fig. 4D) is assigned exclusively to C–P bonding due to the exclusion of C–O bonding (fig. S11B, XPS spectrum) and C–H bonding (table S1). The apparent excitonic state and σ* state in the C *K*-edge XAS (Fig. 4D and supplementary text S6) indicate that the graphite lattice remains a well-ordered planar structure after the synthesis of BP-G. If a P atom is connected to graphite perpendicularly by π bonding, the π* absorption feature (285.4 eV) (Fig. 4D) should become broad, or even split. These analyses suggest the

BP and graphite are mostly connected by BP-G σ bonds at the edge of the basal plane, which is consistent with the high-resolution TEM studies (Fig. 1E).

Our DFT calculation (materials and methods) suggests that the BP-G bonding is an exothermic process with a formation energy of −2.01 eV per P–C bond (Fig. 4E). Notably, our calculation also shows that the energy barrier for Li^+ migration into BP through a graphite–BP boundary is 0.16 eV (Fig. 4F), which is considerably lower than that for Li^+ migration across a reconstructed BP edge (0.52 eV) (fig. S12). The BP-G bonding is thus essential for retaining open channels for Li^+ entry into the BP flakes, leading to an improved D_{Li} for high-rate operation. By physically mixing BP with graphite or Ketjen Black (KB) and a subsequent PANI coating, we prepared two more control materials, (BP/G)/PANI and (BP/KB)/PANI, which showed inferior Li storage performance (fig. S13 and supplementary text S7) to that of the covalently bonded (BP-G)/PANI. The D_{Li} of the control samples is also about two orders of magnitude lower than that of (BP-G)/PANI. The physical mixing of graphitized carbon and BP does not create sufficient quantities of P–C bonds to restrain BP edge reconstruction, and this leads to sluggish charge transfer across the BP edges.

The role of PANI

To reveal the function of PANI, we conducted time-of-flight–secondary ion mass spectrometry (ToF-SIMS) studies on pristine and cycled (BP-G)/PANI and BP-G electrodes to visualize the chemical evolution occurring at the electrode–electrolyte interface (materials and methods and fig. S14). The N-containing secondary ion fragments ($^{13}CN^-$ and C_2N^-) are attributed to PANI, the P-containing species (CP_2^-, C_5P^-, and P_2^-) to BP, and the C_6^- to graphite (fig. S15). The uncycled (BP-G)/PANI electrode shows a considerably stronger signal for N-containing species (fig. S15, A and B) and a much weaker signal for P-containing species (fig. S15, C to E) and C_6^- (fig. S15F) than BP-G, demonstrating the PANI coating on (BP-G)/PANI. Upon cycling, the signals that stand for the key components of SEI (LiO^-, CHO^-, and $LiCO_3^-$) are intensified (Fig. 5, A to C), whereas the signals for P-containing species (Fig. 5C and fig. S16, A and B), which stand for electrode materials, are attenuated. The SEI signals of (BP-G)/PANI are apparently lower than those of BP-G after the same number of cycles (Fig. 5, A to C), indicating inhibited SEI growth on (BP-G)/PANI.

To probe the SEI evolution upon cycling, we collected ToF-SIMS depth profiles from the cycled BP-G and (BP-G)/PANI electrodes (Fig. 5, D to G, and fig. S17). After the first cycle, sharp peak signals of LiO^-, $LiCO_3^-$, and LiF^- (which separately denote Li oxides, carbonates, and fluorides) were detected at the lowest sputter depth (Fig. 5, D and F, and fig. S17, A and C) for both electrodes, suggesting SEI formation. After 100 cycles, the LiO^-, LiF^-, and $LiCO_3^-$ signals remained at a peak at the electrode surface of (BP-G)/PANI, yet they were found all the way into the bulk of the BP-G electrode (Fig. 5, E and G, and fig. S17, B and D), suggesting deep penetration of poorly conducting SEI species (e.g., Li fluorides and carbonates).

These studies show that the PANI coating ensures a stable SEI and prevents continued buildup of poorly conducting species, which may be largely attributable to protection and mediation by a thin layer of PANI swollen with electrolyte (Fig. 5H). The swollen PANI is directly evidenced by the increase in thickness of a PANI disc after immersion into the electrolyte (fig. S18) and is supported by the ToF-SIMS depth profiles of the (BP-G)/PANI electrode at different cycling stages (fig. S16, C and D, fig. S19, and supplementary text S8). The swollen PANI also enables doping of the polymer matrix with Li^+ and protons and uptake of corrosive hydrogen fluoride (fig. S16, E and F, and supplementary text S9) to facilitate charge transport throughout the electrode (Fig. 5I, fig. S20, and supplementary text S10) (*16*, *34*). The inhibited formation of Li fluorides and carbonates and the increased conductivity of swollen PANI facilitate charge transfer at the electrode–electrolyte interface (*35*). Without the PANI coating, the Li fluorides and carbonates build up appreciably (Fig. 5, G and J), degrading the charge transport. Benefiting from the optimized interface design, the (BP-G)/PANI composites retain a stable structure upon cycling, with little change in morphology, electrode thickness, and elemental distributions in the cycled electrode (fig. S21 and supplementary text S11).

By carefully engineering the BP-G interface and (BP-G)/PANI–electrolyte interface, we show that the charge (electron and ion) transport in the composite electrode can be optimized to ensure efficient utilization of the lithium storage capacity of BP under an ultrahigh rate, delivering a combination of high rate, high capacity, and robust cycling performance.

REFERENCES AND NOTES

1. J. B. Goodenough, K.-S. Park, *J. Am. Chem. Soc.* **135**, 1167–1176 (2013).
2. Y. Gogotsi, P. Simon, *Science* **334**, 917–918 (2011).
3. Y. Liu, G. Zhou, K. Liu, Y. Cui, *Acc. Chem. Res.* **50**, 2895–2905 (2017).
4. M. Ko *et al.*, *Nat. Energy* **1**, 16113 (2016).
5. E. J. Berg, C. Villevieille, D. Streich, S. Trabesinger, P. Novák, *J. Electrochem. Soc.* **162**, A2468–A2475 (2015).
6. C. Zhu, R. E. Usiskin, Y. Yu, J. Maier, *Science* **358**, eaao2808 (2017).
7. C.-M. Park, H.-J. Sohn, *Adv. Mater.* **19**, 2465–2468 (2007).
8. X. Ling, H. Wang, S. Huang, F. Xia, M. S. Dresselhaus, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4523–4530 (2015).
9. M. S. Whittingham, *Proc. IEEE* **100**, 1518–1534 (2012).
10. J. Sun *et al.*, *Nano Lett.* **14**, 4573–4580 (2014).
11. J. H. Ryu, J. W. Kim, Y.-E. Sung, S. M. Oh, *Electrochem. Solid-State Lett.* **7**, A306–A309 (2004).
12. W. Li, Y. Yang, G. Zhang, Y.-W. Zhang, *Nano Lett.* **15**, 1691–1697 (2015).
13. Q. Zhang, W. Zhang, W. Wan, Y. Cui, E. Wang, *Nano Lett.* **10**, 3243–3249 (2010).
14. Y. Lee *et al.*, *J. Phys. D Appl. Phys.* **50**, 084003 (2017).
15. J. Wu, N. Mao, L. Xie, H. Xu, J. Zhang, *Angew. Chem. Int. Ed.* **54**, 2366–2369 (2015).
16. G. Ćirić-Marjanović, *Synth. Met.* **177**, 1–47 (2013).
17. M. S. Dresselhaus, A. Jorio, R. Saito, *Annu. Rev. Condens. Matter Phys.* **1**, 89–108 (2010).
18. H. S. Lipson, A. R. Stokes, *Proc. R. Soc. A* **181**, 101–105 (1942).
19. Z. Zhang *et al.*, *Science* **357**, 788–792 (2017).
20. M. Mayo, K. J. Griffith, C. J. Pickard, A. J. Morris, *Chem. Mater.* **28**, 2011–2021 (2016).
21. K. J. Griffith, K. M. Wiaderek, G. Cibin, L. E. Marbella, C. P. Grey, *Nature* **559**, 556–563 (2018).
22. H. Sun *et al.*, *Science* **356**, 599–604 (2017).
23. S. Choi, T.-W. Kwon, A. Coskun, J. W. Choi, *Science* **357**, 279–283 (2017).
24. Y. Li *et al.*, *Nat. Energy* **1**, 15029 (2016).
25. Y. M. Chen, X. Y. Yu, Z. Li, U. Paik, X. W. Lou, *Sci. Adv.* **2**, e1600021 (2016).
26. J. Billaud, F. Bouville, T. Magrini, C. Villevieille, A. R. Studart, *Nat. Energy* **1**, 16097 (2016).
27. P. E. Blanchard, A. P. Grosvenor, R. G. Cavell, A. Mar, *J. Mater. Chem.* **19**, 6015–6022 (2009).
28. W. Hönle, H. G. von Schnering, *Z. Kristallogr. Cryst. Mater.* **155**, 307–314 (1981).
29. Y. Dong, F. J. DiSalvo, *Acta Crystallogr. Sect. E Struct. Rep. Online* **63**, i97–i98 (2007).
30. H. B. Ribeiro *et al.*, *Nat. Commun.* **7**, 12191 (2016).
31. K. Nakanishi, T. Ohta, *J. Phys. Condens. Matter* **21**, 104214 (2009).
32. R. G. Cavell, A. Jürgensen, *J. Electron Spectrosc. Relat. Phenom.* **101-103**, 125–129 (1999).
33. S. Shikata *et al.*, *Appl. Phys. Lett.* **110**, 072106 (2017).
34. M. Wan, J. Yang, *J. Appl. Polym. Sci.* **55**, 399–405 (1995).
35. T. Liu *et al.*, *Nat. Nanotechnol.* **14**, 50–56 (2019).

ACKNOWLEDGMENTS

The authors thank L.-J. Wan from the Institute of Chemistry, CAS, and Y. Luo from USTC for insightful discussion regarding the electrochemistry and phase evolution of BP, and also K.-C. Jiang from Jiangsu TAFEL New Energy Technology for sharing his industry viewpoint about BP. **Funding:** He.J. appreciates funding support from the Natural Science Foundation of China (51672262, 21975243, and 51761145046). S.X. acknowledges financial support from the National Key R&D Program of China (2019YFA0705600). C.C. acknowledges financial support from MOST project 104-2112-M-032-005-MY2. X.W. acknowledges financial support from MOST (2016YFA0200602 and 2018YFA0208603), NSFC (21573204), and Anhui Initiative in Quantum Information Technologies. J.B.G. acknowledges the Lawrence Berkeley National Laboratory BMR Program (grant 7223523). **Author contributions:** He.J. and X.D. designed the research. He.J. and Ho.J. conceived the experiments. Ho.J. conducted the materials preparation. Ho.J., S.X., J.Z., and Z.Q. conducted the electrochemical measurements and/or contributed to data analysis. Ho.J., C.C., W.Y., Y.-R.L., and T.-S.C. conducted the XAS experiments and data analysis. S.X. and W.L. conducted ToF-SIMS experiments and data analysis. X.W., H.W., and Y.W. performed the DFT calculations. H.X. and T.Z. provided black phosphorus materials. He.J., X.D., S.X., Ho.J., and J.B.G. wrote the paper, and all authors were involved in revising the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/192/suppl/DC1
Materials and Methods
Supplementary Text S1 to S11
Figs. S1 to S21
Tables S1 to S4
References (*36*–*50*)

30 September 2018; resubmitted 27 March 2019
Accepted 6 August 2020
10.1126/science.aav5842

QUANTUM GRAVITY

Spacetime from bits

Mark Van Raamsdonk

In the anti–de Sitter/conformal field theory approach to quantum gravity, the spacetime geometry and gravitational physics of states in some quantum theory of gravity are encoded in the quantum states of an ordinary nongravitational system. Here, I demonstrate that this nongravitational system can be replaced with an arbitrarily large collection of noninteracting systems ("bits") placed in a highly entangled state. This construction makes manifest the idea that spacetime geometry emerges from entanglement between the fundamental degrees of freedom of quantum gravity and that removing this entanglement is tantamount to disintegrating spacetime. This setup also reveals that the entangled states encoding spacetimes may be well represented by a certain type of tensor network in which the individual tensors are associated with states of small numbers of bits.

In the anti–de Sitter/conformal field theory (AdS-CFT) or "holographic" approach to quantum gravity (*1*), spacetime physics and gravitational dynamics are encoded in the states of a nongravitational quantum system, often a strongly interacting quantum field theory with invariance under scaling transformations, that is, a conformal field theory. Over the past decade, there has been increasing evidence that quantum entanglement between the degrees of freedom in this nongravitational system plays a crucial role in the emergence of spacetime [see, for example, (*2*–*7*), or (*8*) for a review]. One notable example of this phenomenon is that placing two distinct CFTs without interactions between them into a particular entangled state amounts to connecting the two encoded spacetimes via a wormhole (*2*). Partly on the basis of this example, it was argued that (*5*, *6*) for any spacetime described by a CFT state, the geometry is "built up" by the entanglement present in the CFT state and that removing this entanglement destroys the encoded spacetime. However, the continuous nature of the CFT systems that encode the spacetimes makes this argument challenging. In the usual examples, part of the spacetime structure (the fixed asymptotic behavior) is directly related to the continuous geometrical space upon which the CFT is defined. Further, the local degrees of freedom of the CFT interact strongly with those around them, and completely disentangling the various parts would require an infinite amount of energy.

In this paper, I introduce a framework for holography in which the fundamental degrees of freedom are a large collection of elementary systems that do not interact with one another and have no intrinsic spatial arrangement. The individual systems are conformal field theories on spaces with boundaries; these can be thought of as "bits" of the original CFT. I show that states of a holographic CFT can be replaced by entangled states of these bits so that the system still describes a single connected spacetime that is arbitrarily close to the one encoded in the original CFT state. This construction makes it clear that it is possible to build up very generic connected spacetimes by entangling discrete noninteracting systems; because entanglement is the only thing that relates these systems, it is also manifest that the encoded spacetime completely disintegrates when the systems are disentangled.

In the bits framework, I show that there is a natural way to represent the state with arbitrary precision using a type of tensor network. This representation confirms previous work suggesting that certain tensor network states of discrete elementary subsystems (*4*, *9*, *10*) capture the qualitative features of the entanglement structure of CFT states encoding spacetimes.

Boundary conformal field theory bits

The motivation for this construction is the idea of cutting up a holographic conformal field theory into a large number of noninteracting pieces but putting the pieces into a quantum state that approximates a state of the original CFT encoding some spacetime. To make the construction precise, it is necessary to define what is meant by a CFT living on a piece of space with a boundary. Generally speaking, the boundary conditions for the fields at the edges need to be described. For any CFT, there are various choices of boundary conditions. A special subset of these conditions preserve the scaling symmetry and conformal invariance (*11*) and define what is known as a boundary conformal field theory, or BCFT [see, for example, (*12*)].

A set of "BCFT bits" associated with a CFT on some spatial geometry M is a collection of BCFTs defined on a set of "sanded" pieces $\tilde{M}_i$ of M. Here, $\{\tilde{M}_i\}$ are defined by cutting M into a set of simply connected pieces $\{M_i\}$ and "sanding the edges," that is, $\tilde{M}_i$ is a large subset of the interior of M_i with a smooth boundary (Fig. 1). Each BCFT is defined from the original CFT with the same choice of boundary conditions, so the BCFT-bit system is specified by the choice of $\{\tilde{M}_i\}$ and the choice of boundary condition.

The goal here is to consider some state of the CFT that corresponds to a smooth geometry and to approximate this state by a particular entangled state of the discretized BCFT-bit system. I will argue that the newly generated state still encodes a smooth connected geometry that is closely related to the original geometry.

Entangling BCFT bits via the Euclidean path integral

The Euclidean path integral gives a mechanism by which to define a quantum state of a CFT on a spatial geometry M given a spatial geometry H of one higher dimension with boundary M (e.g., Fig. 2A). The quantum state is specified by assigning a probability amplitude to each possible configuration of fields on M. This probability amplitude is given by the average of a specific quantity (the exponential of the negative of the Euclidean action) over

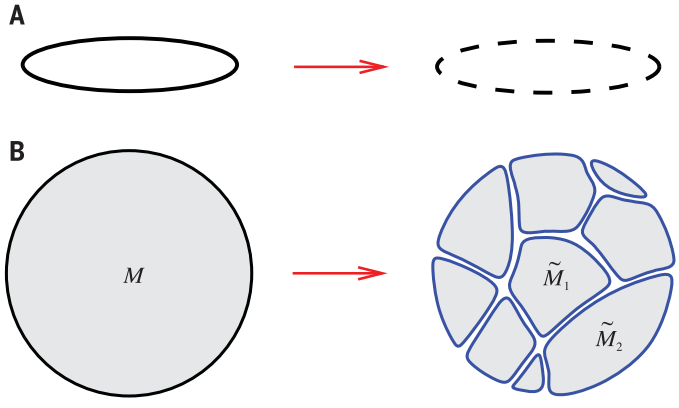


Fig. 1. BCFT bits. Starting with a conformal field theory on some connected space M, the associated BCFT bits $\tilde{M}_i$ are defined as a collection of conformal field theories with boundaries living on spatial geometries that approximate a set of pieces of the original space. This is illustrated for the case where M is (**A**) a circle or (**B**) a two-dimensional (2D) sphere.

Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada. **Email: mav@phas.ubc.ca**

all possible field configurations on H subject to the condition that the fields take the specified values on M. For holographic CFTs, this construction gives a natural way to define states that encode dual spacetimes with a good classical description [see, for example, (*13*–*15*) and references therein]. The spacetime being described can be changed by varying the interior geometry of H and adding sources for various CFT operators on the interior of H; for example, arbitrary linearized perturbations around empty anti–de Sitter spacetime can be obtained by choosing the right geometry and sources (*15*).

Similarly, it is also possible to define the state of a BCFT on a spatial geometry M with boundary B given a higher-dimensional geometry H with boundary $M \cup G$, where G is also bounded by B, as shown in Fig. 2C. In this case, the boundary conditions at G are taken to be those of the BCFT under consideration.

The Euclidean path integral can be used to define natural entangled states of noninteracting CFTs or BCFTs by taking the H to be a connected geometry with a disconnected boundary M (*16*). For example, taking H to be a finite cylinder, M will be a pair of circles, and the path integral on H defines an entangled state of a pair of CFTs encoding two spacetimes connected by a wormhole (Fig. 2B). I will use this idea to define an entangled state of BCFT bits by taking H to be a connected geometry whose boundary is the collection $\{\tilde{M}_i\}$.

More specifically, I would like to define a state of the BCFT bits that is related to some state of the original CFT on M defined by the path integral over H. To do so, I define a geometry $\tilde{H}$ obtained from H by removing smooth "grooves" at the surface so that the part of the boundary remaining is $\tilde{M}_i$. This is depicted in Fig. 3 for a 1+1-dimensional CFT defined on a spatial circle. The boundary of $\tilde{H}$ is $\{\cup_i \tilde{M}_i\} \cup G$, where G corresponds to the surface of the grooves. The state of the BCFT bits can now be defined by performing the Euclidean path integral for the BCFT over the geometry $\tilde{H}$, with the appropriate boundary conditions imposed at G.

The state $|\Psi_{\tilde{H}}\rangle$ and the possible dual geometry will depend on the details of $\tilde{H}$. However, I will now argue that if I have been sufficiently gentle with my carpentry tools—that is, if $\{\tilde{M}_i\}$ is close enough to $\{M_i\}$, and if $\tilde{H}$ is close enough to H—that the state $|\Psi_{\tilde{H}}\rangle$ of the BCFT bits encodes a geometry that is closely related to the original geometry described by the CFT state $|\Psi_H\rangle$.

Spacetimes dual to Euclidean path integral states

Let us now recall how to understand the spacetime geometries encoded by holographic CFT states described by a Euclidean path integral. According to the AdS/CFT correspondence, features of the encoded geometry can be deduced from the CFT by evaluating the expectation value of various local and nonlocal observables in the CFT state. For the states in this work, these expectation values are computed using a path integral over a surface $\bar{H}H$ that is obtained by gluing H to its mirror image $\bar{H}$ along M, as shown in Fig. 4, C and I (*17*). The operators of interest are inserted along the junction. The AdS/CFT correspondence implies that this CFT path integral is equal to a gravitational path integral whose average over geometries is dominated by a single Euclidean geometry X_H, obtained by solving the gravitational equations with boundary conditions that X_H is asymptotically anti–de Sitter with boundary geometry $\bar{H}H$ (*18*).

The Lorentzian spacetime geometry associated with my CFT state is simply related to the Euclidean geometry X_H. The geometry X_H has a reflection symmetry inherited from the geometry of $\bar{H}H$ (Fig. 4, C and I, depicts the geometry X_H sliced along the plane of symmetry). The surface left invariant under this symmetry has a geometry $(X_H)_0$. To find the spacetime associated with this state, I use the geometry $(X_H)_0$ (and the condition that time derivatives of fields vanish here) as the initial conditions for the real-time gravitational equations. The solution is a spacetime X_H^L corresponding to the state. More formally, X_H^L can be understood as an analytic continuation of X_H. The CFT state at time $(t) = 0$ strictly encodes only the region of this spacetime that is spacelike separated from the $t = 0$ slice at the boundary (*19*), given that the spacetime outside this region can be altered by changes to the CFT Hamiltonian before or after $t = 0$.

The crux of my subsequent argument is that, despite the considerable differences between the original CFT and the collection of BCFT bits as physical systems, the geometry $\tilde{H}$ used to define the BCFT-bit state is a small perturbation to the geometry H used to define the CFT state. Thus, it might be expected that the procedure I have just outlined gives rise to a spacetime dual to the BCFT bits that is almost the same as the spacetime dual to the CFT state.

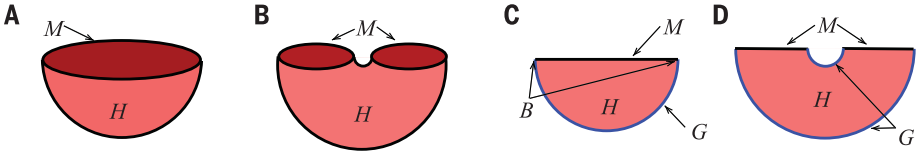


Fig. 2. Euclidean path integrals. The integrals defining (**A**) a state of a 2D CFT on a circle, (**B**) an entangled state of two 2D CFTs each on a spatial circle, (**C**) a state of a 2D BCFT on an interval, and (**D**) an entangled state of two 2D BCFTs.

Fig. 3. Entangled state of many BCFTs. The path integral defining some state of a holographic CFT (left) is approximated by a path integral defining an entangled state of many BCFTs (right). The second path integral geometry is obtained from the first one by introducing additional small boundary components (shown in blue).

Geometry of the BCFT-bit states

I would like to understand how the Euclidean gravity solution corresponding to a BCFT on $\tilde{\bar{H}}\tilde{H}$ differs from that corresponding to the CFT on $\bar{H}H$ (Fig. 4, C versus I). The main obstacle here is understanding how the presence of a boundary in $\tilde{\bar{H}}\tilde{H}$ (geometrically described as the space G glued to a mirror image of itself along B) affects the gravity calculation. This question was considered originally in (*20*) and later in more detail in (*21*). As discussed in those papers, if the BCFT state has a geometrical dual $X_{\tilde{H}}$, this dual must itself have a boundary component in addition to the asymptotically AdS boundary with boundary geometry $\tilde{H}$. In detailed examples using ultraviolet (UV) complete theories of gravity, this boundary component corresponds to a place where some compact internal dimensions smoothly degenerate (*22*, *23*), similar to how a Kaluza-Klein circle shrinks to zero in a bubble-of-nothing spacetime. I expect that the specific way such a boundary is realized will depend on the boundary conditions chosen for the BCFT.

As a simple model, the suggestion in (*20*, *21*) can be used to introduce an explicit end-of-the-world (ETW) brane with constant tension and Neumann boundary conditions. In that case, as long as a given boundary region of $\tilde{H}$ is small compared with other geometrical

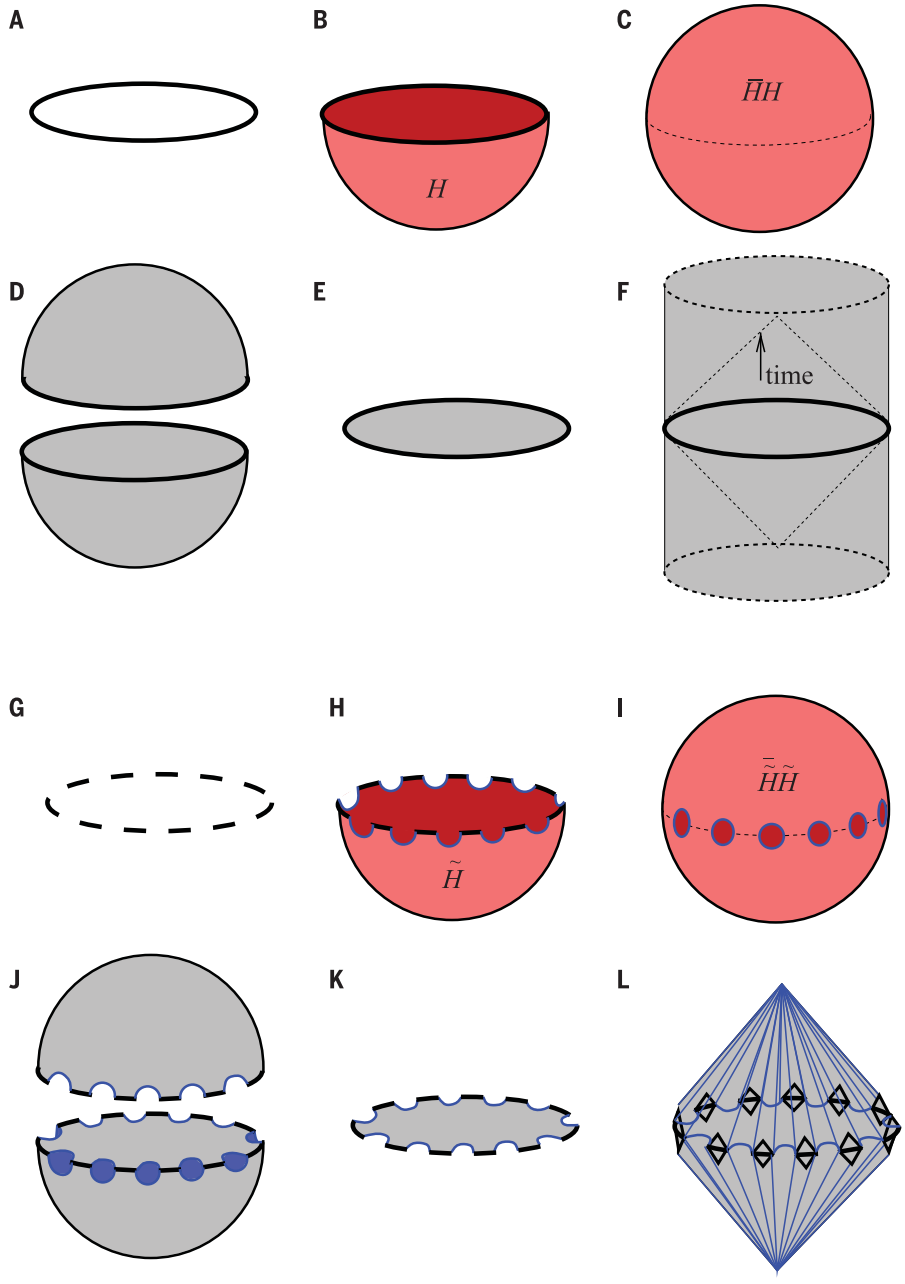


Fig. 4. Lorentzian geometries from path-integral states. (**A**) A CFT on a circle. (**B**) Euclidean path integral defining a holographic CFT state and (**C**) the path integral used to compute observables in this state. Operators can be inserted on the dashed line. (**D**) Euclidean gravity solution corresponding to the path integral in (C). (**E**) Spatial slice at the time-symmetric point serves as the initial data for the Lorentzian solution. (**F**) Lorentzian solution associated with our state. The interior of the causal diamond (dashed lines) is the part encoded by the CFT state at $t = 0$. (**G** to **L**) Equivalent constructions for the BCFT-bit states. Each BCFT bit is a boundary CFT on an interval.

features of $\tilde{H}$ (such as the distance to other boundary components) the ETW brane ending on that boundary component stays localized to the vicinity of that boundary component, as depicted in Fig. 4, J and K (blue surfaces). The ETW brane may source bulk fields and affect the rest of the geometry, but for a fixed location away from the ETW brane, these effects will become negligible in the limit where the boundary component is taken to be small.

For 1+1-dimensional CFTs, a more direct argument can be given that does not rely on a particular holographic model for BCFTs. In this case, $\tilde{\bar{H}}\tilde{H}$ is a two-dimensional Euclidean geometry, as in the example of Fig. 4I. Here, each component of G can be taken to be a small circle, as shown in Fig. 4I. If the circle is sufficiently small relative to the distance to other boundary components and operator insertions, it is expected that (as in the operator product expansion) its insertion is equivalent to the insertion of a sum of local CFT operators. In the limit where the circle becomes small, this operator sum approaches the identity operator (*24*), so the insertion of the boundary G has a very small effect.

According to these arguments, the Euclidean geometry associated with $\tilde{\bar{H}}\tilde{H}$ should be almost the same as that associated with $\bar{H}H$. But this is not quite true for the corresponding Lorentzian geometries. The reason is that no matter how small the boundary components of $\tilde{\bar{H}}\tilde{H}$ are, they still change the asymptotic geometry of the slice that serves as the initial data for the Lorentzian evolution (Fig. 4J). The result is that in the Lorentzian picture, the initial data slice is almost the same as that for the original CFT state, but because of the differences in asymptotics at the boundary, there may be some type of shockwave evolving forward and backward from each introduced boundary component. In the ETW brane picture, this shockwave can be understood as a Lorentzian ETW brane whose worldvolume is part of a hyperboloid (the analytic continuation of the hemispherical ETW branes in the Euclidean picture of Fig. 4J). The Lorentzian spacetime is depicted in Fig. 4L. In a limit with very many BCFT bits and very many small boundary components, this shockwave or ETW brane will propagate outward from the full asymptotic boundary of the initial data slice. Thus, the BCFT-bit version of a holographic state faithfully reproduces the region of the original spacetime whose points are spacelike separated from the $t = 0$ slice of the boundary.

A necessary consequence of these shockwaves is that it is no longer possible to move causally between asymptotic regions associated to different BCFT bits; such causal paths would imply interactions between the bits that are not present in this study's setup. In this sense, the modified spacetime has causal properties similar to a two-sided black hole, where it is not possible to move between the two asymptotic regions. As with a black hole, the spacetime considered here may also contain a future singularity, which might be avoided by reintroducing some interactions between neighboring BCFT bits.

Emergence of the radial direction and quantum error correction

It is of interest what part of the dual spacetime can be reconstructed having access only to density matrices for subsystems of a number (n) of adjacent BCFT bits. In general, the

region encoded by a subsystem has been shown (*25*–*28*) to be the region bounded by the minimal-area surface enclosing a portion of the geometry that includes the subsystem A at the boundary; the areas of such surfaces give the entropies of the corresponding subsystems (*3*). As shown in Fig. 5, the region for a subsystem of n adjacent bits extends farther into the bulk as n is increased; standard calculations (*3*) show that the proper distance into the bulk increases as the logarithm of n (*29*). To learn about regions deeper within the geometry, knowledge of entanglement structure and correlations at longer scales is needed. This relation between the geometrical depth and the scale of entanglement aligns with the well-known connection between the radial direction in the geometry and renormalization group flow in the field theory: physics deeper within the bulk geometry corresponds to physics further into the infrared in the dual quantum theory (*1*, *30*). The setup in this study also manifests the recently explained feature (*31*) that the encoding of geometry into holographic systems is qualitatively similar to a quantum error-correcting code: Local physics near the point P in Fig. 5 can be recovered with knowledge of the state of bits 1 to 6 or bits 4 to 9, but it cannot be recovered with knowledge of only bits 4 to 6, that is, the intersection of these systems. Thus, the information about physics near point P is encoded nonlocally and redundantly in the BCFT bits in the way that a logical qubit is encoded nonlocally in the physical qubits of a quantum error-correcting code.

Although this section has focused on the encoding of the spatial geometry, I emphasize that the construction described in the previous section leads to a full spacetime dual to the BCFT-bit state. However, it is an interesting open question how the emergent bulk time and, more generally, the local bulk physics at various spacetime points can be understood from the BCFT perspective.

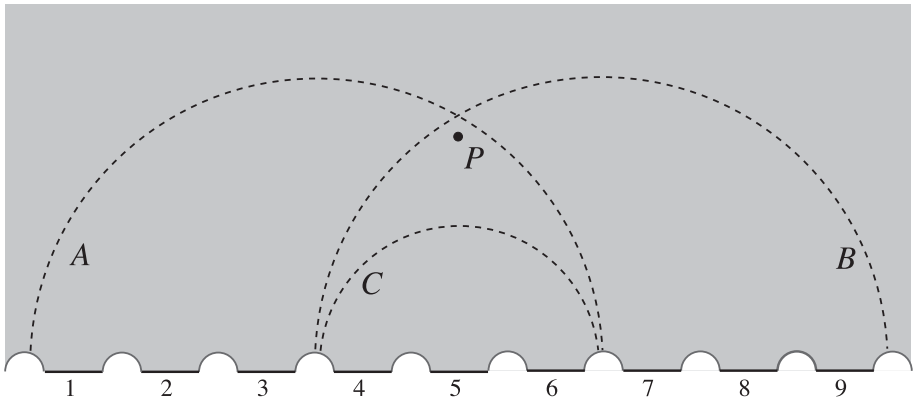


Fig. 5. Reconstructing spacetime from adjacent bits. The density matrices for bits 1 to 6, 4 to 9, and 4 to 6 encode the geometry of regions A, B, and C, respectively, bounded by the minimal area extremal surfaces enclosing these bits. Physics at P can be recovered from bits 1 to 6 or bits 4 to 9 but not from bits 4 to 6.

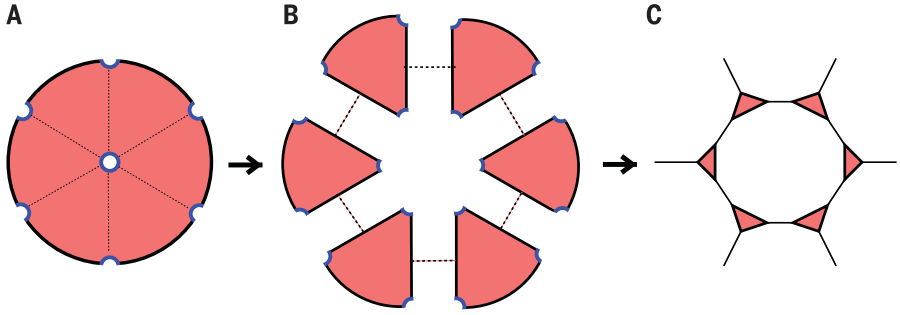


Fig. 6. Tensor network representation. (**A**) An extra circular boundary component is added to the interior of $\tilde{H}$ in the path integral for a six-BCFT-bit state. (**B**) The path integral can now be decomposed into a product of path integrals defining three BCFT-bit states. The field configurations for the BCFT bits connected by dashed lines are equated and integrated over the path integral equivalent of the pair projection that connects a tensor network. (**C**) The tensor network representation of the state.

Tensor networks for holographic BCFT-bit states

The BCFT-bit construction of holographic states bears a close resemblance to tensor network toy models of holography (*4*, *9*, *10*) in that there are explicit multipart entangled states of a discrete set of noninteracting constituents. I will now demonstrate that, as in the toy models, the states considered herein may be represented arbitrarily well by a type of tensor network, where the tensors correspond to states of small numbers of BCFT bits. The construction involves additional small changes to $\tilde{H}$.

I have argued previously that the introduction of a small boundary component to $\tilde{H}$ has a vanishingly small effect on the dual Euclidean geometry in the limit that the size goes to zero, except very close to the insertion locus. Thus, if additional boundary components are introduced away from M, these should have a negligible effect on the BCFT-bit state as the newly introduced boundary components are taken to be infinitesimal.

By introducing these extra boundaries (for example, by drilling narrow holes through $\tilde{H}$), it is possible to represent the newly fabricated state as a tensor network, by decomposing the path integral into chunks (Fig. 6). Here, each individual tensor corresponds to the state of a small number of BCFT bits, which are the original BCFT bits for the external legs or newly constructed internal BCFT bits. The internal edges of the tensor network correspond in path integral language to identifying the field configurations on the two BCFT bits joined by the edge and integrating over these. In quantum language, this corresponds to projecting the state of the pair of BCFT bits onto a maximally entangled state. The full set of these projections joins up the tensor network, defining what is known as a projected entangled pair state [see, for example, (*32*)].

There is a great deal of freedom in how to build up a tensor network representation: The placement of the boundary components can be chosen, as can the surfaces along which to break up $\tilde{H}$ to give the internal BCFT bits. So, as expected, it is possible to have many tensor networks that represent the same state. The networks here apparently have a closer connection to the geometry of the path integral defining the state rather than the geometry of the space being encoded; such a connection was emphasized recently in (*33*) [see also

(*34*, *35*)]. However, other recent work (*36*, *37*) suggests that by a particular optimization of the path-integral geometry over the geometries related by conformal invariance, the path-integral geometry actually becomes the geometry of the bulk spatial slice. Similarly, it may be that some optimization of the geometrical slicing produces a tensor network similar to those (*9*, *10*) where the network geometry has a close connection to the bulk spatial geometry.

The limit of many BCFT bits

For simplicity, the various examples shown in the figures involve only a small number of BCFT bits. In this case, the individual BCFT bits might carry information about a substantial portion of the dual geometry. However, the faithful reproduction of the original spacetime applies equally well when the number of BCFT bits is taken to be very large, so long as the modifications leading from H to $\tilde{H}$ are kept small (e.g., each $\tilde{M}_i$ is still a large subset of M_i). In the limit where the BCFT bits are all small compared with any scale associated with M or the original CFT state, I expect that the individual bits carry almost no information about the geometry being represented by the collection of BCFT bits, apart from the asymptotic behavior of the bulk fields at a single boundary location corresponding to the bit. Thus, the spacetime geometry is almost entirely encoded in the entanglement structure of a multipart BCFT-bit system. In this system, it is manifestly true that disentangling the bits causes the corresponding spacetime to disintegrate, as suggested for continuous CFTs (*6*). I emphasize that in the BCFT-bit system, the individual bits have no intrinsic location relative to one another, so it is not only the radial direction of the spacetime that "emerges."

A universal entanglement–gravity duality?

Because the entanglement structure is playing such a key role here, and because the individual bits carry almost no information about the interior of the spacetime, it is of interest what properties of the BCFT bit are really required here. If the BCFT bits are replaced with some other type of bit but the entanglement structure is kept the same, does this still encode the same spacetime and gravitational physics? This might seem too optimistic; the specific theory of gravity encoded in a CFT state has a certain number of dimensions and a certain set of fields in addition to the metric, and these are related to the dimensionality and the operator content of the CFT. If an attempt is made to replace BCFT bits from one CFT with BCFT bits from another CFT, perhaps of a different dimensionality, it may seem impossible that the same gravitational physics is being described, given that the new CFT may correspond to a completely different theory of gravity.

But there is a way around this obstacle: The optimistic scenario could work if there is fundamentally only one theory of quantum gravity. This is in line with expectations from string theory (*38*), where the study of string dualities suggests that various UV complete gravitational theories in various dimensions can be understood as descending through compactification and dualities from 11-dimensional M-theory. Thus, even if the gravitational physics in one region of spacetime has the fields and interactions associated with a particular low-energy theory of gravity, the fields and interactions in another region of spacetime could correspond to a different low-energy theory if there is some transition region in between where the properties of the compact extra dimensions change.

In the example where BCFT bits from one CFT are replaced with BCFT bits associated with another CFT, the asymptotically AdS regions very close to the BCFT bits (near the diamonds in Fig. 4L) will clearly have different physics. But moving inward in the radial directions, there may be a transition (e.g., with some compact dimensions changing shape or topology) such that inside a radial position associated with a boundary scale containing a large number of BCFT bits, the physics is that of the spacetime described by the original CFT. If the BCFT bits are replaced with more general quantum systems (e.g., collections of qubits or macaroni), it may be that the asymptotic region no longer has a geometrical description, but the same interior region emerges.

The idea that it is possible to obtain a precise description of quantum gravitational physics starting from sufficiently many copies of an arbitrary quantum system is intriguing but certainly does not follow from any of the arguments in this paper. Nevertheless, it is fascinating that the possible unity of gravitational theories as suggested by string theory leaves open the possibility of such an exact and universal entanglement–gravity duality.

REFERENCES AND NOTES

1. J. M. Maldacena, *Adv. Theor. Math. Phys* **2**, 231–252 (1998).
2. J. M. Maldacena, *J. High Energy Phys.* **2003**, 021 (2003).
3. S. Ryu, T. Takayanagi, *Phys. Rev. Lett.* **96**, 181602 (2006).
4. B. Swingle, *Phys. Rev. D* **86**, 065007 (2012).
5. M. Van Raamsdonk, arxiv.org/abs/0907.2939 [hep-th] (17 July 2009).
6. M. Van Raamsdonk, *Gen. Relativ. Gravit.* **42**, 2323–2329 (2010).
7. J. Maldacena, L. Susskind, *Fortschr. Phys.* **61**, 781–811 (2013).
8. M. Van Raamsdonk, in *TASI 2015: New Frontiers in Fields and Strings: Proceedings of the 2015 Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, 1–26 June 2015*, J. Polchinski, P. Vieira, O. DeWolfe, Eds. (World Scientific, 2017), pp. 297–351.
9. F. Pastawski, B. Yoshida, D. Harlow, J. Preskill, *J. High Energy Phys.* **2015**, 149 (2015).
10. P. Hayden *et al.*, *J. High Energy Phys.* **2016**, 9 (2016).
11. For example, the vacuum state of the CFT on a half space with these boundary conditions at the edge preserves an SO(d-1,2) of the SO(d,2) conformal symmetry.
12. J. Cardy, arXiv:hep-th/0411189v2 [hep-th] (20 February 2008).
13. K. Skenderis, B. C. van Rees, *Phys. Rev. Lett.* **101**, 081601 (2008).
14. M. Botta-Cantcheff, P. J. Martínez, G. A. Silva, *J. High Energy Phys.* **2016**, 171 (2016).
15. D. Marolf, O. Parrikar, C. Rabideau, A. I. Rad, M. Van Raamsdonk, *J. High Energy Phys.* **2018**, 77 (2018).
16. V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, S. F. Ross, *Class. Quantum Gravity* **31**, 185015 (2014).
17. Any complex sources on H should be conjugated in $\tilde{H}$, but I will restrict to the case of real sources; the resulting geometry will then have a time-reversal symmetry.
18. The asymptotic behavior of other fields in the geometry is fixed by the sources for the corresponding operator in the path-integral action.
19. This is the domain of dependence of the region $(X_H)_0$.
20. A. Karch, L. Randall, *J. High Energy Phys.* **2001**, 063 (2001).
21. T. Takayanagi, *Phys. Rev. Lett.* **107**, 101602 (2011).
22. M. Chiodaroli, E. D'Hoker, M. Gutperle, *J. High Energy Phys.* **2012**, 5 (2012).
23. M. Chiodaroli, E. D'Hoker, M. Gutperle, *J. High Energy Phys.* **2012**, 177 (2012).
24. To understand this limit, consider inserting the circular boundary in a disk path integral. The resulting geometry is conformally equivalent to a finite cylinder; in the limit where the circle becomes small, this cylinder becomes infinitely long, so the path integral gives the vacuum state. In this limit, the inserted boundary is equivalent to inserting the identity operator.
25. B. Czech, J. L. Karczmarek, F. Nogueira, M. Van Raamsdonk, *Class. Quantum Gravity* **29**, 155009 (2012).
26. M. Headrick, V. E. Hubeny, A. Lawrence, M. Rangamani, *J. High Energy Phys.* **2014**, 162 (2014).
27. D. L. Jafferis, A. Lewkowycz, J. Maldacena, S. J. Suh, *J. High Energy Phys.* **2016**, 4 (2016).
28. X. Dong, D. Harlow, A. C. Wall, *Phys. Rev. Lett.* **117**, 021601 (2016).
29. An interesting point is that the areas and entropies are finite here, because they end on ETW branes rather than at the asymptotic boundary of the spacetime.
30. L. Susskind, E. Witten, arxiv.org/abs/hep-th/9805114 [hep-th] (19 May 1998).
31. A. Almheiri, X. Dong, D. Harlow, *J. High Energy Phys.* **2015**, 163 (2015).
32. R. Orús, *Ann. Phys.* **349**, 117–158 (2014).
33. A. Milsted, G. Vidal, arxiv.org/abs/1807.02501 [cond-mat.str-el] (6 July 2018).
34. R. Kōnig, V. B. Scholz, *Phys. Rev. Lett.* **117**, 121601 (2016).
35. L. Y. Hung, W. Li, C. M. Melby-Thompson, *J. High Energy Phys.* **2019**, 170 (2019).
36. P. Caputa, N. Kundu, M. Miyaji, T. Takayanagi, K. Watanabe, *Phys. Rev. Lett.* **119**, 071602 (2017).
37. P. Caputa, N. Kundu, M. Miyaji, T. Takayanagi, K. Watanabe, *J. High Energy Phys.* **2017**, 97 (2017).
38. J. Polchinski, *String Theory. Volume II: Superstring Theory and Beyond* (Cambridge Monographs on Mathematical Physics, Cambridge Univ. Press, 2007).

ACKNOWLEDGMENTS

Funding: This work was supported in part by a Simons Investigator award, by the Simons Foundation grant "It from Qubit: Simons Collaboration on Quantum Fields, Gravity and Information" and by the Natural Sciences and Engineering Research Council of Canada. **Competing interests:** The author has no competing interests.

31 July 2019; accepted 8 August 2020
10.1126/science.aay9560

CORONAVIRUS

In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges

Beata Turoňová[1,2]*, Mateusz Sikora[3]*, Christoph Schürmann[4]*, Wim J. H. Hagen[1], Sonja Welsch[5], Florian E. C. Blanc[3], Sören von Bülow[3], Michael Gecht[3], Katrin Bagola[6], Cindy Hörner[4,7], Ger van Zandbergen[6,8,9], Jonathan Landry[10], Nayara Trevisan Doimo de Azevedo[10], Shyamal Mosalaganti[1,2], Andre Schwarz[1], Roberto Covino[3,11], Michael D. Mühlebach[4,7], Gerhard Hummer[3,12]†, Jacomine Krijnse Locker[13]†, Martin Beck[1,2]†

The spike protein (S) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is required for cell entry and is the primary focus for vaccine development. In this study, we combined cryo–electron tomography, subtomogram averaging, and molecular dynamics simulations to structurally analyze S in situ. Compared with the recombinant S, the viral S was more heavily glycosylated and occurred mostly in the closed prefusion conformation. We show that the stalk domain of S contains three hinges, giving the head unexpected orientational freedom. We propose that the hinges allow S to scan the host cell surface, shielded from antibodies by an extensive glycan coat. The structure of native S contributes to our understanding of SARS-CoV-2 infection and potentially to the development of safe vaccines.

The spike surface protein (S) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is required to initiate infection (*1*). It binds to the angiotensin-converting enzyme 2 (ACE2) (*2*, *3*) to mediate viral entry. S also determines tissue and cell tropism. Mutations may alter the host range of the virus and enable the virus to cross species barriers (*4*, *5*). Vaccine efforts focus on neutralizing antibodies that block infection by binding to S.

S is a trimeric class I viral fusion protein (*6*) with a club-like shape of ~20 nm in length. The ectodomain consists of a head, which has been extensively studied in vitro. It is connected to the membrane by a slender stalk. The three receptor binding domains (RBDs) of the S head are conformationally variable, which may relate to receptor binding. In the closed conformation, the RBDs are shielded by the N-terminal domains (NTDs). In the open conformation, one RBD is exposed upward away from the viral membrane (*2*, *3*). Previous studies resolved roughly two-thirds of the predicted 22 N-linked glycans that are thought to shield S against antibodies (*2*, *3*). It remains unknown whether the distribution of the conformational states and the glycosylation pattern observed with recombinant protein in vitro are representative of the native state generated during viral assembly. Furthermore, little is known about the stalk of S and how its conformational variability within the virion may affect the accessibility of epitopes for neutralizing antibodies and facilitate viral entry.

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Meyerhofstr. 1, 69117 Heidelberg, Germany. [2]Department of Molecular Sociology, Max Planck Institute of Biophysics, Max-von-Laue Str. 3, 60438 Frankfurt am Main, Germany. [3]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue Str. 3, 60438 Frankfurt am Main, Germany. [4]Division of Veterinary Medicine, Paul Ehrlich Institute, Paul Ehrlich Strasse 51-59, 63225 Langen, Germany. [5]Central Electron Microscopy Facility, Max Planck Institute of Biophysics, Max-von-Laue Str. 3, 60438 Frankfurt am Main, Germany. [6]Division of Immunology, Paul Ehrlich Institute, Paul Ehrlich Strasse 51-59, 63225 Langen, Germany. [7]German Center for Infection Research, Gießen-Marburg-Langen, Germany. [8]Institute for Immunology, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany. [9]Research Center for Immunotherapy (FZI), University Medical Center, Johannes Gutenberg-University Mainz, Mainz, Germany. [10]Genomics Core Facility, EMBL, Meyerhofstr. 1, 69117 Heidelberg, Germany. [11]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany. [12]Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany. [13]Electron Microscopy of Pathogens Unit, Paul Ehrlich Institute, Paul Ehrlich Strasse 51-59, 63225 Langen, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: gerhard.hummer@biophys.mpg.de (G.H.); jacomina.krijnselocker@pei.de (J.K.L.); martin.beck@biophys.mpg.de (M.B.)

SARS-CoV-2 virions present prefusion S in an irregular pattern

To structurally analyze SARS-CoV-2 S in situ, we passaged the virus through tissue culture cells and used sucrose centrifugation to purify it from the inactivated supernatant (see materials and methods). We acquired a large-scale cryo–electron tomography dataset that consists of 266 tilt series covering >1000 viruses. Visual inspection of the tomographic reconstructions revealed a very-high-quality data set in which individual protein domains were clearly visible (Fig. 1A and movie S1). On average, 40 copies of the S trimer resided on the surface. S proteins appeared to be distributed randomly on the viral surface without any significant tendency to cluster (Fig. 2A).

S was mostly present in the prefusion conformation (Fig. 1B). Postfusion conformations (*7*, *8*) were very rare (<0.1%), which appears typical for Vero E6 host cells (*9*). Sanger sequencing and immunoblot analysis revealed that the furin site for proteolytic cleavage into the S1 and S2 fragments (*5*, *10*) was lost during tissue culture passage (Fig. 1C and fig. S1), consistent with previous studies (*11*, *12*). However, the isolate contained the Asp^{614}→Gly (D614G) allele (*13*, *14*). Large-scale sequencing of RNA isolated from tissue culture supernatant confirmed both findings (supplementary materials).

Subtomogram averaging with NovaSTA (*15*) and STOPGAP (*16*) resulted in a cryo–electron microscopy (cryo-EM) map of the S head at 7.9 Å resolution (fig. S2), in which secondary structure elements and individual glycosylation sites were clearly discernible (Fig. 2, B and C). Classification suggested that about half of S was present in the fully closed conformation. A considerable fraction of the remaining subtomograms had one RBD exposed (fig. S3). Structural analysis of the asymmetric unit yielded an average map of the closed conformation at an overall resolution of 4.9 Å. In particular, the cluster of parallel helices in the center of the head was clearly resolved (Fig. 2D and fig. S4).

By contrast, the stalk connecting the S head to the viral membrane appeared to be dynamic. Although the head was fully contained in the tomographic map, only the top of the stalk domain was resolved. Emerging from the neck of the spike head, it contains an 11-residue Leu repeat sequence (L1141, L1145, and L1152) and adopts an unusual right-handed coiled coil, consistent with a recent single-particle structure of the S head (*7*). We will henceforth refer to this part of the stalk domain as the "upper leg." Right-handed trimeric coiled coils were long thought to be absent from the structural proteome (*17*) but can be seen in the post-fusion structure of S from the related mouse hepatitis virus (*18*).

A stalk with three flexible hinges connects S to the viral membrane

The tomographic images suggest the presence of flexible hinges in the stalk. Stalks of individual S proteins are clearly visible in the tomograms (Fig. 1B), but, after averaging, their density declined sharply at the end of the trimeric coiled coil that forms the upper leg (Fig. 2B). Moreover, the head exhibited large positional and orientational freedom. It was tilted up to ~90° with respect to the normal at distances of 5 to 35 nm from the membrane (Fig. 2E). We grouped our subtomograms into four classes, according to their distance from the bilayer, and averaged them separately. At an intermediate distance, parts of the stalk and bilayer were resolved, suggesting a more defined conformation (Fig. 2F). We then subselected ~3200 particles in which the head was oriented roughly perpendicular to the membrane. In the resulting average, the stalk domain was resolved (fig. S5A). Visual inspection of the respective subtomograms, in which the stalk domains are clearly observed, further corroborated the idea of a kinked stalk with

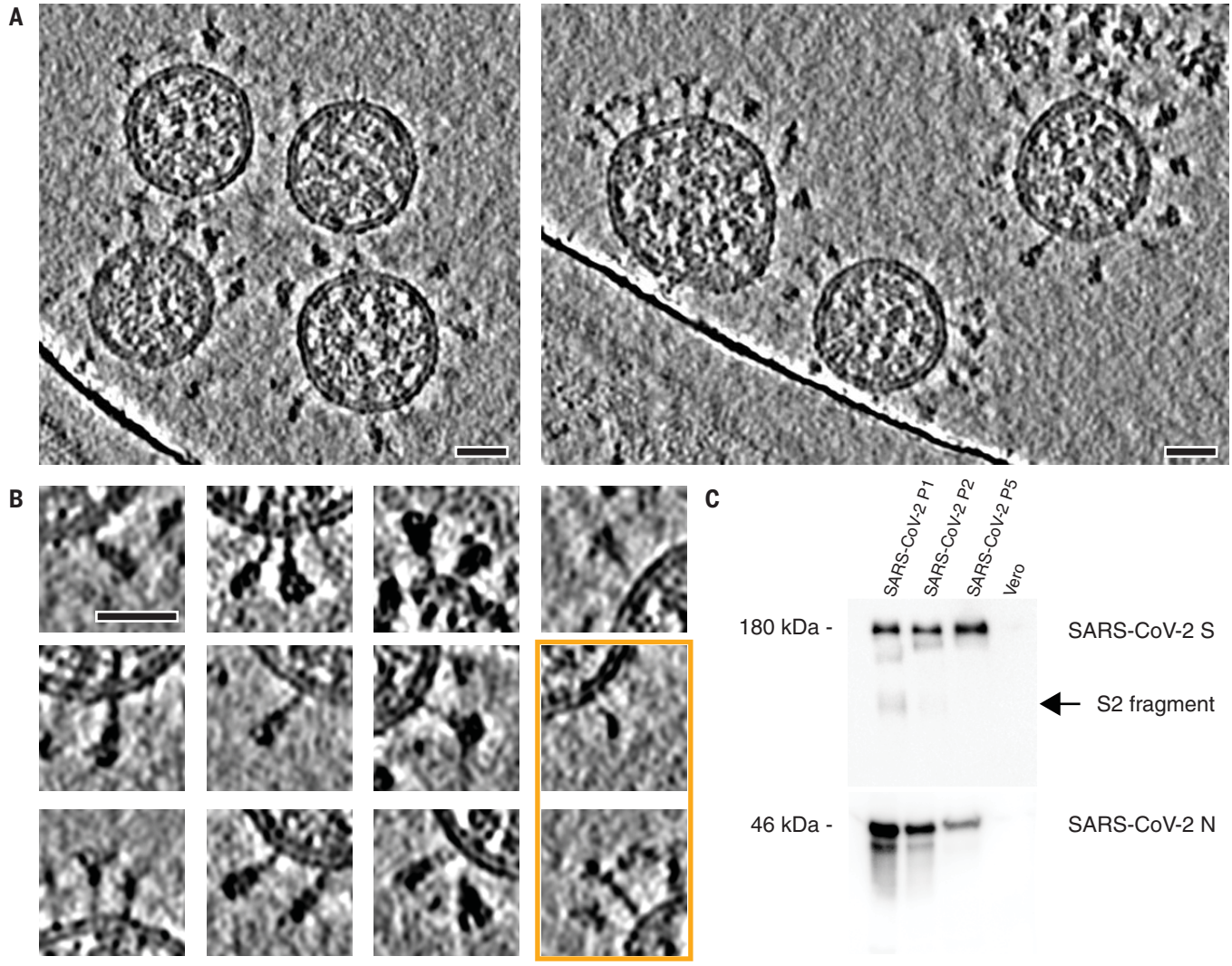


Fig. 1. Cryo–electron tomography of SARS-CoV-2 virions. (**A**) Slices through tomographic reconstructions of SARS-CoV-2 virions. Scale bars, 30 nm. (**B**) Same as (A), but tomograms are arranged as a gallery to highlight specific features of S. All domains, including the transmembrane part, are clearly resolved. Although most of S is reminiscent of the prefusion conformation, the particles framed in orange resemble the postfusion conformation as described by Cai *et al*. (*7*). Scale bar, 30 nm. (**C**) Immunoblot showing the loss of cleavage products of SARS-CoV-2 S with uncleaved S (180 kDa) remaining, within five passages (P1 to P5) through tissue culture (loading control using anti-N antibody). N, nucleocapsid protein.

potentially several hinges (Fig. 2F). Local refinement of the lower part of the stalk (henceforth referred to as the "lower leg") resulted in a moderately resolved structure that would be consistent with the continuation of the coiled coil below a flexible hinge (henceforth referred to as the "knee") (fig. S5B).

Molecular dynamics (MD) simulations helped us to pinpoint the molecular origins of the flexibility seen in the tomograms. We performed a 2.5-μs-long all-atom MD simulation of a 4.1 million atom system containing four glycosylated S proteins anchored into a patch of viral membrane and embedded in aqueous solvent (Fig. 3A). In the simulations, the S heads remained stable. The stalks, however, exhibited pronounced hinging motions at the junctions between the S head and the upper leg ("hip"), between the upper and lower legs ("knee"), and between the lower leg and the transmembrane domain ("ankle"). This observation was consistent with discrete leg segments seen in the raw tomograms (Fig. 3, B and C). The hip joint flexed the least (16.5° ± 8.8°), followed by the ankle (23.0° ± 11.7°) and the knee (28.4° ± 10.2°) (Fig. 3D and fig. S6). However, the limited sampling in the MD simulation may not have covered the full range of motions (compare Fig. 2E and fig. S6D).

Structures of S seen along the MD trajectory fit well into the tomographic density of S proteins protruding from the viral surface (Fig. 4A). In particular, the joints of the hip, knee, and ankle of the MD snapshots aligned with kinks in the density visualized by cryo-EM. For a more detailed view, we flexibly fitted suitable snapshots of the MD simulations into the subtomogram averages classified according to the distance of the head from the membrane (compare Fig. 2F to Fig. 4B). Hinge bending gives the S stalk the flexibility required to connect heavily tilted S heads to the viral membrane.

As a result of hinge bending, the stalk is diluted out in subtomogram averages focused on the head (Figs. 2, B, C, and F, and 4B). Stalks were visible if the heads were aligned with the membrane normal (fig. S5A) or if the stalks themselves were averaged separately (fig. S5B). To test this interpretation, we calculated the electron density averaged over the entire MD trajectory with aligned S heads. Filtered to a comparable resolution, this calculated 3D map was highly similar to the subtomogram averages (Fig. 4B). In rare cases, the coiled coil near the membrane appears to be unfolded in the original tomograms

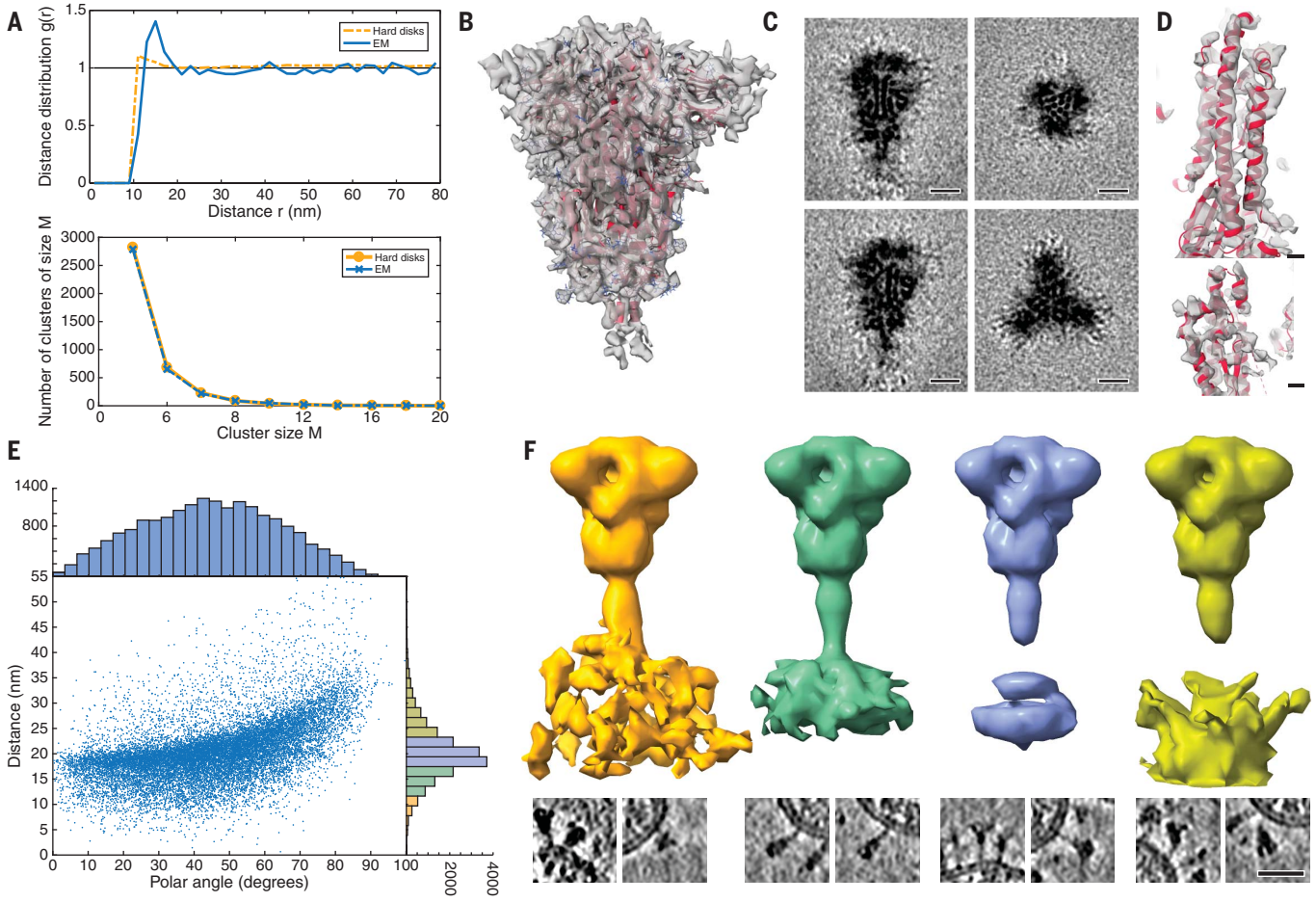


Fig. 2. Subtomogram analysis of SARS-CoV-2 S protein. (**A**) Distance (top) and cluster-size distributions (bottom) of S on the viral surface, with nonoverlapping hard disks of 10-nm diameter as a reference. (**B**) Subtomogram average of the ectodomain of S, shown isosurface rendered and fitted with the previously published atomic model as determined by single-particle EM (PDB ID 6VXX). Transparent gray, subtomogram average; red, secondary structure elements; blue, glycosylation sites. (**C**) Same subtomogram average but shown as slices through the reconstruction. Scale bars, 5 nm. (**D**) Detail of the average of the symmetric unit of S. Scale bars, 5 Å. (**E**) Distribution of the angular orientation and distance of the ectodomain with respect to the bilayer. (**F**) On the basis of the initial subtomogram averaging, positions of the spike head were classified according to their distance from the lipid bilayer (supplementary materials). (Top) Averages of the resulting classes are shown isosurface rendered; distance increases from left to right. (Bottom) Examples of individual particles are shown as slices. At an optimal distance, the stalk domain stretches out and is resolved. Scale bar, 30 nm.

(fig. S5C) and continuous with the disordered loops of the MD model.

Extensive N-glycosylation covers the surface of S

The predicted N-glycosylation sites, many already annotated in single-particle EM maps (*2*), were generally very pronounced in the subtomogram averages. The electron density of N-glycans averaged over the MD trajectory was highly consistent with the tomographic map (Fig. 5A). Clustered glycosylation sites were visible in the raw density before averaging (e.g., protruding from the lower part of the S head; Fig. 5B). Our analysis of individual sites in subtomogram averages further supports the notion that the spikes were decorated with rather bulky glycan chains (Fig. 5C). Notably, a number of sequons were resolved with more-pronounced branching than previously reported (*19*). By contrast, the two predicted O-glycosylation sites (*20*) lacked excess density (fig. S7A). Sequon N17LT, owing to its location on the unstructured N terminus, was not localized in the density (fig. S7B), but elongated features protruding from the tip of the N-terminal domain (fig. S7B) suggested the presence of sequons N74GT and N149KS.

N-glycosylation is also predicted on the knee (N1158HT and N1173AS) and the ankle (N1194ES) in regions not previously resolved by single-particle EM (Fig. 3A). We observed that these positions generally appeared bulkier in tomographic reconstructions than one might expect if they were not glycosylated (Figs. 1B and 5D). Additional density was very clearly observed in subtomogram averages (Fig. 5, E and F, and fig. S5, A and B), and consistent electron density calculated from the MD trajectory aligned on the lower leg (fig. S7C). N-glycosylation in this region of S might protect the functionally important hinges from antibody binding and help to keep them flexible.

Discussion

The two primary structural analysis techniques combined in this study are complementary. Our MD simulations revealed three flexible hinges (hip, knee, and ankle) within the stalk, consistent with the tomographic data. One might speculate that the high degree of conformational freedom of S on the viral surface is important for the mechanical robustness of the virus or may facilitate motions that interfere with antibody access to the stalk. It might also allow S to engage the relatively flat surface of host cells with higher avidity (Fig. 5, G and H). Future tomographic studies of actual infection events might further explore these topics. In contrast to the prefusion conformation of S, the postfusion conformation previously observed in vitro and in situ (*7*, *9*), as well as in this study (Fig. 1B), is apparently inflexible. To the best of our knowledge, extensive flexibility comparable to that of the prefusion S stalk has not been reported for other class I viral fusion proteins,

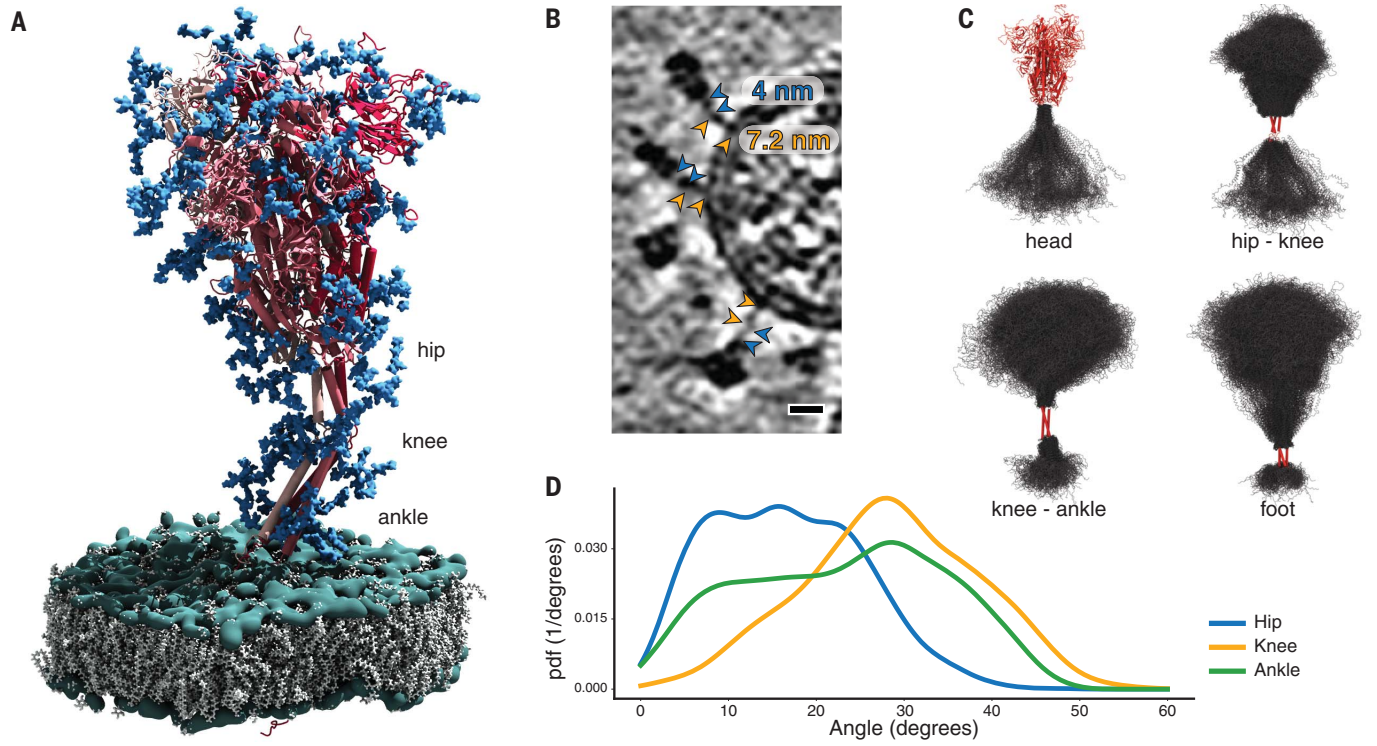


Fig. 3. MD simulations of SARS-CoV-2 S protein. (**A**) Model of the S protein. The three individual chains of S are shown in shades of red, N-glycosylation in blue, lipids of the endoplasmic reticulum–like membrane in gray, and phosphates in green. "Hip," "knee," and "ankle" mark positions of the three flexible hinges. (**B**) Examples of the hinges as seen in the deconvoluted tomograms. Blue and orange arrowheads indicate the upper and lower legs, respectively, with their typical lengths indicated. Scale bar, 10 nm. (**C**) Hinge flexibility in the MD simulation illustrated through backbone traces (gray) at 75-ns intervals with different parts of the S protein fixed (red). (**D**) Probability density functions (pdf) for hinge bending angles at the hip, knee, and ankle.

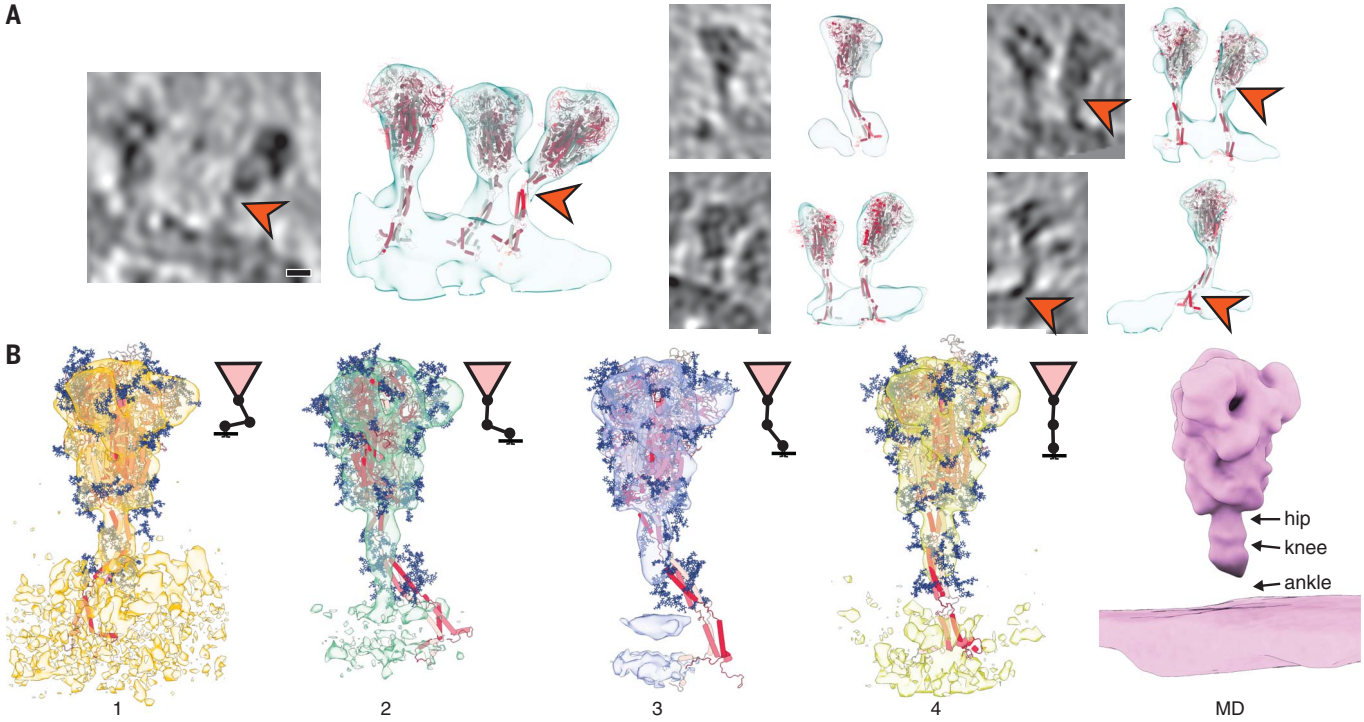


Fig. 4. Fitting of molecular simulations into cryo–electron tomograms. (**A**) Slices through tomograms (left) and isosurface-rendered tomograms with snapshots of the respective MD simulations superimposed without flexible fitting (right). The hinges of the stalk domain predicted by structural modeling (orange arrowheads) are consistent with the tomographic data. Scale bar, 5 nm. (**B**) Fit of snapshots of MD simulations into the classes obtained for different distances of the head from the membrane (1 to 4), as presented in Fig. 2F. Shorter distances are concomitant with a stronger bending of the hinges and a lateral displacement of the stalk. Average MD density filtered to a resolution comparable to the subtomogram averages is shown as an isosurface rendering (right).

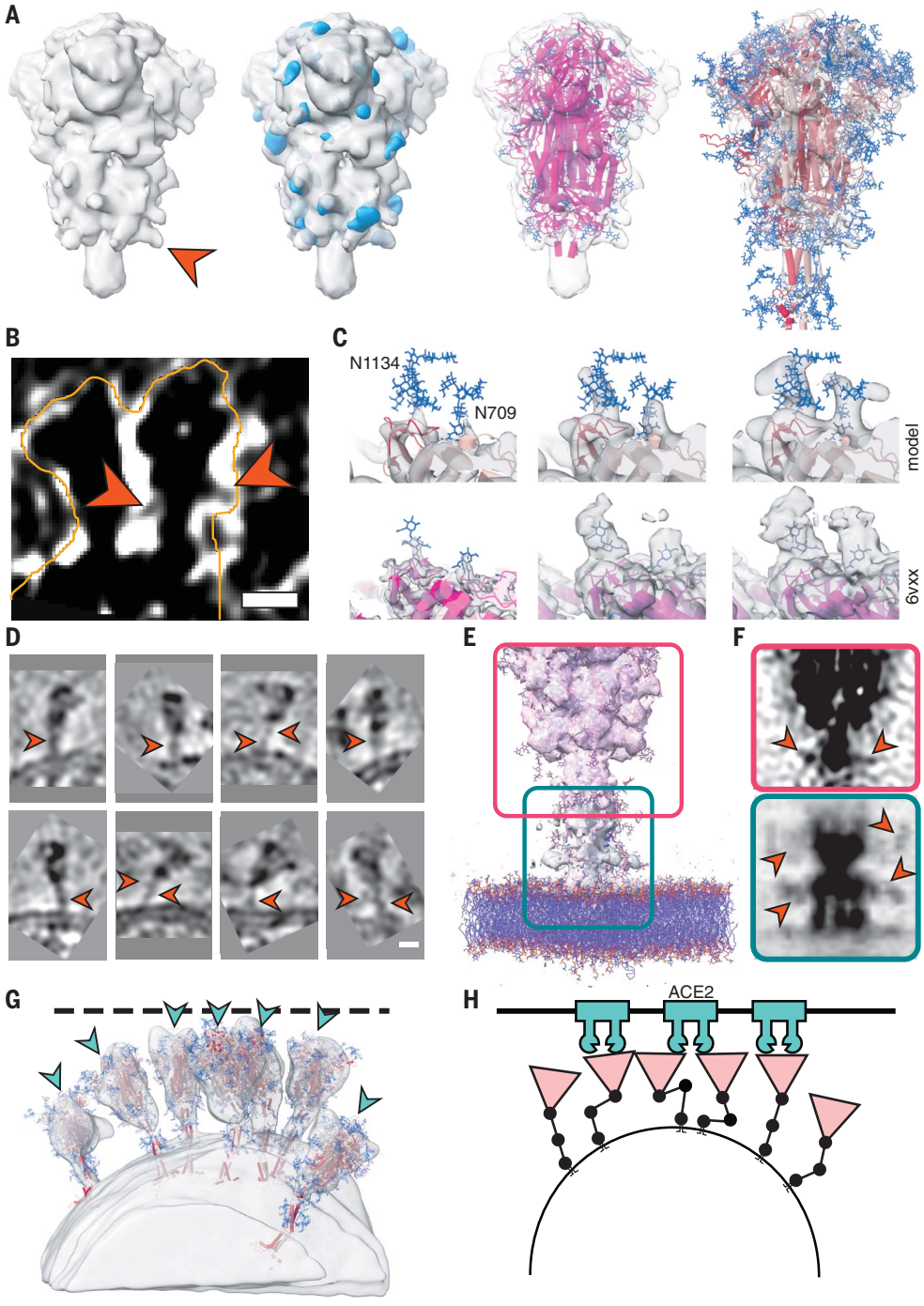


Fig. 5. Analysis of S protein glycosylation sites and epitopes. (**A**) N-glycosylation sites are clearly discernible in the subtomogram average of the head. From left to right: Isosurface rendering of subtomogram average with an individual N-glycosylation site indicated (orange arrowhead); superimposed with the MD-calculated density for all annotated N-glycosylation sites; superimposed with previous structural model of the head (PDB ID 6VXX); and superimposed with a snapshot of the MD simulations. N-glycosylation sites are shown in blue. (**B**) Tomographic slice highlighting an N-glycosylation site (orange arrowheads) in the original data. Scale bar, 5 nm. (**C**) Highlight of N-glycosylation positions 709 and 1134 of the MD simulations (top) and in a previous structural model (bottom; PDB ID 6VXX, EMDB 21452). The subtomogram average is shown superimposed at different isosurface thresholds (transparent gray). Extensive additional density is visible. (**D** to **F**) The stalk domain is heavily glycosylated at the hinges. (D) Example tomographic slices with bulky density at the hinge positions (orange arrowheads). Scale bar, 5 nm. (E) Superposition of the subtomogram averages (transparent gray isosurfaces) of the head (framed red) and the stalk domain (framed green), with a respective snapshot of the MD simulations emphasizing the glycosylation at the hinges. (F) Same as (E) but shown as a maximum intensity projection through the subtomogram averages. Orange arrowheads indicate bulky density at hinges. (**G**) Fits of snapshots from MD simulations into the surface of a virion. The tomogram is shown isosurface rendered in transparent gray. The position of epitopes for neutralizing antibodies at the RBDs are indicated with blue arrowheads. (**H**) Cartoon illustrating a hypothetical docking event in which the hinges facilitate the engagement of multiple instances of S with their receptors.

including HIV env, influenza HA, or Ebola GP. However, influenza HA attaches to micelles with a short linker permitting up to 25° bending (*21*).

A particularly unusual feature masked at the edge of the resolved density of single-particle structures but well resolved in the subtomogram averages is the short right-handed coiled coil at the top of the prefusion stalk. Because this feature is lost in the postfusion structure as resolved for SARS-CoV (*8*), we speculate that it is only marginally stable, priming the protein for a large structural reorganization in a spring-loaded viral fusion mechanism. Indeed, all three hinges are disassembled in the transition to the postfusion conformation and placed outside the structural core (*7*, *8*).

Overall, the observed distribution of S on the surface of the virion and its conformers is highly consistent with the findings of other studies (*9*, *22*, *23*). Host cell-type–dependent differences in the abundance of pre- and postfusion conformation (*9*, *22*) may depend on different levels of ACE2 and the serine protease TMPRSS2 (*10*). Whether the furin cleavage site plays a role here remains to be addressed. A notable difference is the higher abundance of S on the viral surface observed in this study compared with others (*22*, *23*).

The fully closed prefusion conformation of S was abundant in situ. This finding emphasizes that the highly engineered, recombinant versions of S locked into this conformation (*24*, *25*) may be valuable tools for vaccine development, although there are also differences to the in situ structure. N-glycosylation sites appeared very bulky in the tomographic map compared with previous single-particle analysis, suggesting that decoration with sugars

may be more extensive on S assembled in infected cells than on S expressed recombinantly. Our map is suggestive of additional N-glycosylation at the hinges of the stalk domain and possibly on the tips of the S NTDs. The native glycosylation pattern defines the accessibility of epitopes on the crowded viral surface (*19*), where the NTD and stalk domains appear occluded by neighboring spikes (Fig. 5G). A lack of excess density at the predicted O-glycosylation sites indicates that N-glycosylation dominates.

By using cryo–electron tomography of intact viruses, we were able to resolve functionally important parts of S, including its connection to the viral membrane and its glycan coat, which were masked in studies of recombinant detergent-solubilized protein. Beyond S, our large-scale tomographic dataset contains rich, high-resolution structural information on SARS-CoV-2 particles in their native context. The in situ structures of several key viral components—including the nucleocapsid and the M protein that is highly enriched in the membrane—remain enigmatic. Our data might thus be explored to resolve such features in the future. Furthermore, high-resolution structural models can be fitted directly into the tomographic reconstructions, emphasizing the high quality of the data. This strategy might thus help us to build structural models of entire virions.

REFERENCES AND NOTES

1. C. A. M. de Haan, P. J. M. Rottier, *Adv. Virus Res.* **64**, 165–230 (2005).
2. A. C. Walls *et al*., *Cell* **181**, 281–292.e6 (2020).
3. D. Wrapp *et al*., *Science* **367**, 1260–1263 (2020).
4. R. J. Hulswit, C. A. de Haan, B.-J. Bosch, *Adv. Virus Res.* **96**, 29–57 (2016).
5. F. Li, *Annu. Rev. Virol.* **3**, 237–261 (2016).
6. S. C. Harrison, *Virology* **479–480**, 498–507 (2015).
7. Y. Cai *et al*., *Science* 10.1126/science.abd4251 (2020).
8. S. Duquerroy, A. Vigouroux, P. J. Rottier, F. A. Rey, B. J. Bosch, *Virology* **335**, 276–285 (2005).
9. S. Klein *et al*., bioRxiv 16706 [Preprint]. 16 August 2020. https://doi.org/10.1101/2020.06.23.167064.
10. S. Belouzard, J. K. Millet, B. N. Licitra, G. R. Whittaker, *Viruses* **4**, 1011–1033 (2012).
11. S.-Y. Lau *et al*., *Emerg. Microbes Infect.* **9**, 837–842 (2020).
12. N. S. Ogando *et al*., *J. Gen. Virol.* 10.1099/jgv.0.001453 (2020).
13. C. Rothe *et al*., *N. Engl. J. Med.* **382**, 970–971 (2020).
14. L. Zhang *et al*., bioRxiv 148726 [Preprint]. 12 June 2020. https://doi.org/10.1101/2020.06.12.148726.
15. B. Turoňová, turonova/novaSTA: NovaSTA, Version v1.0, Zenodo (2020); https://doi.org/10.5281/zenodo.3973623.
16. W. Wan, williamnwan/STOPGAP: STOPGAP 0.7.1, Version 0.7.1, Zenodo (2020); https://doi.org/10.5281/zenodo.3973664.
17. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
18. A. C. Walls *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11157–11162 (2017).
19. A. C. Walls *et al*., *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
20. A. Shajahan, N. T. Supekar, A. S. Gleinich, P. Azadi, *Glycobiology* 10.1093/glycob/cwaa042 (2020).
21. D. J. Benton *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10112–10117 (2018).
22. Z. Ke *et al*., *Nature* 10.1038/s41586-020-2665-2 (2020).
23. H. Yao *et al*., *Cell* 10.1016/j.cell.2020.09.018 (2020).
24. R. Henderson *et al*., *Nat. Struct. Mol. Biol.* 10.1038/s41594-020-0479-4 (2020).
25. X. Xiong *et al*., *Nat. Struct. Mol. Biol.* 10.1038/s41594-020-0478-5 (2020).

ACKNOWLEDGMENTS

The cryo–electron tomography data was collected at the EMBL Heidelberg Cryo Electron Microscopy Service Platform. The genome sequencing was done at the Genomics Core Facility of EMBL Heidelberg. We thank EMBL (B.T., W.J.H.H., S.M., A.S., and M.B.) and the Max Planck Society (B.T., M.S., S.W., F.E.C.B., S.v.B., M.G., S.M., R.C., G.H., and M.B.) for support and the Max Planck Computing Data Facility for providing computational resources. B.T. acknowledges W. Wan (Vanderbilt University) for helpful discussions. J.K.L. acknowledges R. Eberle (PEI) for support. R.C. acknowledges the Frankfurt Institute for Advanced Studies for support. The authors are indebted to G. Dobler and R. Wölfel, Bundeswehr Institute for Microbiology, for providing SARS-CoV-2 strain MUC-IMB1. **Funding:** We acknowledge a generous SuperMUC-NG computing allocation at the Leibniz Supercomputing Centre (M.S., S.v.B., M.G., F.E.C.B., R.C., and G.H.), the Human Frontier Science Program (RGP0026/2017; S.v.B. and G.H.), the German Ministry of Health (C.S.), the German Center for Infection Research (C.H. and M.D.M.), and the Loewe Center DRUID from the Justus Liebig University Giessen (J.K.L.) for funding. M.S. acknowledges support from the Austrian Science Fund FWF (Schrödinger Fellowship, J4332-B28). **Author contributions:** B.T.: experimental design, tomographic reconstruction, particle picking, subtomogram averaging, structural analysis, and paper writing. M.S.: modeling design, molecular dynamics simulations, structural analysis, and paper writing. C.S.: experimental design, virus purification, and biochemical analysis and sequencing. W.J.H.H.: experimental design, cryo-EM data acquisition, and tomographic reconstruction. S.W.: experimental design, sample preparation and screening, and data analysis. F.E.C.B., S.v.B., M.G., and R.C.: molecular dynamics simulations and structural analysis. K.B.: experimental design and virus purification. C.H.: experimental design and virus growth. G.v.Z.: experimental design and supervision. J.L.: sequencing. N.T.D.d.A.: sequencing. S.M.: subtomogram averaging. A.S.: tomographic reconstruction and particle picking. M.D.M.: experimental design and supervision. G.H.: modeling design, data analysis, supervision, and paper writing. J.K.L.: experimental design, supervision, and paper writing. M.B.: experimental design, supervision, and paper writing. **Competing interests:** None declared. **Data and materials availability:** The original tilt series have been deposited in the Electron Microscopy Public Image Archive (EMPIAR-10453). Subtomogram averages were deposited in the Electron Microscopy Database under accession numbers EMD-11222 (S-trimer RBDs closed), EMD-11347 (S-trimer with one fully open RBD), and EMD-11223 (asymmetric unit with closed RBD). The viral sequencing reads have been deposited in the European Nucleotide Archive repository under accession ID PRJEB39737. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/203/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 and S2
References (*26*–*60*)
MDAR Reproducibility Checklist
Movie S1

View/request a protocol for this paper from *Bio-protocol*.

26 June 2020; accepted 13 August 2020
Published online 18 August 2020
10.1126/science.abd5223

DEVELOPMENTAL BIOLOGY

The mole genome reveals regulatory rearrangements associated with adaptive intersexuality

Francisca M. Real[1,2], Stefan A. Haas[3], Paolo Franchini[4], Peiwen Xiong[4], Oleg Simakov[5], Heiner Kuhl[6], Robert Schöpflin[1,2], David Heller[3], M-Hossein Moeinzadeh[3], Verena Heinrich[3], Thomas Krannich[3], Annkatrin Bressin[3], Michaela F. Hartmann[7], Stefan A. Wudy[7], Dina K. N. Dechmann[8,9], Alicia Hurtado[10,11], Francisco J. Barrionuevo[10,11], Magdalena Schindler[1,2], Izabela Harabula[1], Marco Osterwalder[12,13], Michael Hiller[14,15,16], Lars Wittler[17], Axel Visel[12,18,19], Bernd Timmermann[1], Axel Meyer[4], Martin Vingron[3], Rafael Jiménez[10,11], Stefan Mundlos[1,2,20]*, Darío G. Lupiáñez[1,2,20,21]*

Linking genomic variation to phenotypical traits remains a major challenge in evolutionary genetics. In this study, we use phylogenomic strategies to investigate a distinctive trait among mammals: the development of masculinizing ovotestes in female moles. By combining a chromosome-scale genome assembly of the Iberian mole, *Talpa occidentalis*, with transcriptomic, epigenetic, and chromatin interaction datasets, we identify rearrangements altering the regulatory landscape of genes with distinct gonadal expression patterns. These include a tandem triplication involving *CYP17A1*, a gene controlling androgen synthesis, and an intrachromosomal inversion involving the pro-testicular growth factor gene *FGF9*, which is heterochronically expressed in mole ovotestes. Transgenic mice with a knock-in mole *CYP17A1* enhancer or overexpressing *FGF9* showed phenotypes recapitulating mole sexual features. Our results highlight how integrative genomic approaches can reveal the phenotypic impact of noncoding sequence changes.

Differences in genome sequence and structure provide the molecular foundation for phenotypic diversity across species and enable environmental adaptation. In evolutionary genetics, linking genomic alterations to phenotypic traits has largely relied on candidate gene (*1*) or linkage-mapping analyses (*2*). However, the combination of next-generation sequencing with proximity-ligation methods, in particular Hi-C, allows the generation of chromosome-scale genome assemblies (*3*), introducing ample possibilities for comparative genomics. Hi-C also enables the integration of three-dimensional (3D) genome structure with transcriptional control. Vertebrate genomes are spatially organized into regulatory units, termed topologically associating domains (TADs) (*4*, *5*). Although TADs are generally preserved across species (*6*, *7*), studies of human disease highlighted that alterations in TAD organization can cause changes in gene expression and developmental phenotypes by rewiring enhancer-promoter contacts (*8*–*10*). Thus, analytical strategies that consider 3D organization and regulatory potential become essential for a comprehensive annotation of genomes. Here, we introduce a phylogenomic strategy that combines comparative whole-genome, epigenomic, transcriptomic, and chromatin interaction data to identify phenotype-associated genomic changes. We demonstrate the power of this approach by elucidating the molecular underpinnings of generalized intersexuality in female moles, an evolutionary trait distinctive among mammals.

In mammals, sex is determined genetically. Genetic elements direct the differentiation of the bipotential gonad into either testicular or ovarian tissue, which, in turn, leads to the development of sex-specific anatomical, hormonal, and behavioral differences (*11*). An exception to this paradigm occurs in moles (family Talpidae), in which XX-genotypic females have an intersex phenotype in at least eight species (*12*, *13*). Although male moles have normal testes, genotypic females develop ovotestes instead of ovaries (Fig. 1A). These unusual gonads are composed of an ovarian part (OP) that fully supports sexual reproduction and a testicular part (TP) that lacks fertile germ cells but contains typical male cell populations, such as androgen-producing Leydig cells (*12*) (fig. S1). As a consequence of increased androgen synthesis, female moles develop masculinized external genitalia, as well as prominent muscles and aggressive behavior (*14*), traits that likely represent adaptations to a subterranean lifestyle.

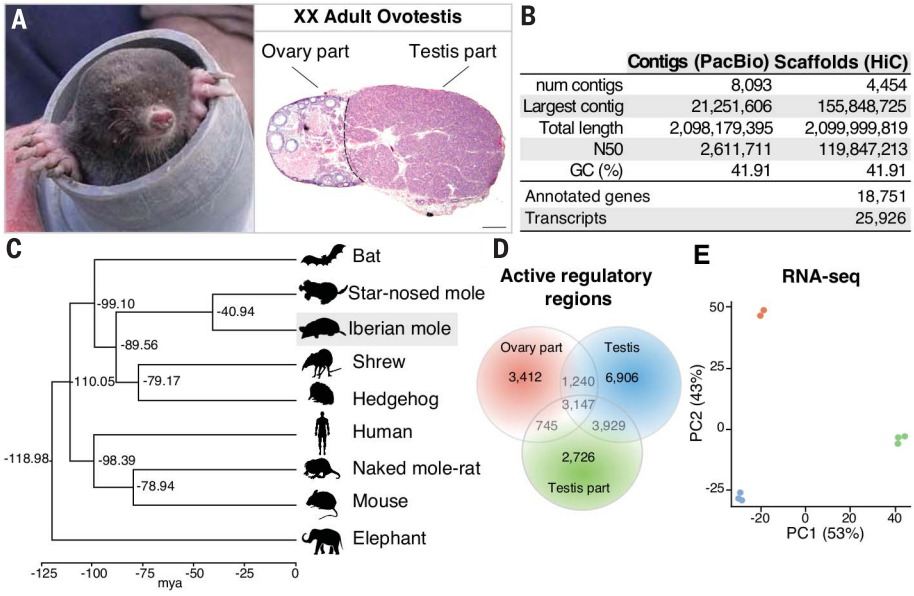
The Iberian mole genome

To investigate the molecular origins of mole ovotestes, we generated a chromosome-scale genome assembly for *Talpa occidentalis* based on long- and short-read sequencing and scaffolded using Hi-C data (Fig. 1B and fig. S2). The assembly comprises 2.099 gigabases, and up to 30% is made of transposable elements, with a repeat profile that differs from that of closely related mammals (fig. S3 and supplementary text). Combining RNA sequencing (RNA-seq) datasets and homology-based predictions, we identified 18,751 genes, including 2370 single-copy orthologs. This gene subset was used to determine 1580 one-to-one orthologous genes in nine species and build a phylogenetic tree (Fig. 1C). Our analysis confirms moles as a distinct family in the order Eulipotyphla, with shrews and hedgehogs being the most closely related species (*15*) (supplementary text).

Epigenetic and transcriptional landscape of mole gonadal development

We generated epigenetic and transcriptomic profiles of mole gonads at 7 days postpartum (P7), processing TPs and OPs separately (fig. S2). Specifically, we produced chromatin immunoprecipitation sequencing (ChIP-seq) datasets against histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) to segment the mole genome into functional states for each tissue. Additionally, we performed assays for transposase-accessible chromatin using sequencing (ATAC-seq) (fig. S2) and intersected both datasets to predict active enhancers in each tissue (22,105 in total) (Fig. 1D and data S1). Although the TP and testis shared a higher number of putative enhancers than the TP and OP, the large number of TP-specific putative

[1]RG Development & Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany. [2]Institute for Medical and Human Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany. [3]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [4]Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, 78457 Konstanz, Germany. [5]Department of Molecular Evolution and Development, University of Vienna, 1090 Vienna, Austria. [6]Department of Ecophysiology and Aquaculture, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [7]Steroid Research & Mass Spectrometry Unit, Laboratory for Translational Hormone Analytics in Paediatric Endocrinology, Division of Paediatric Endocrinology & Diabetology, Center of Child and Adolescent Medicine, Justus Liebig University, Giessen, Germany. [8]Department of Migration and Immuno-Ecology, Max Planck Institute for Animal Behavior, Radolfzell, Germany. [9]Department of Biology, University of Konstanz, Konstanz, Germany. [10]Departamento de Genética, Universidad de Granada, Granada, Spain. [11]Instituto de Biotecnología, Centro de Investigación Biomédica, Universidad de Granada, Armilla, Granada, Spain. [12]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [13]Department for BioMedical Research (DBMR), University of Bern, 3008 Bern, Switzerland. [14]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [15]Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany. [16]Center for Systems Biology Dresden, 01307 Dresden, Germany. [17]Department of Developmental Genetics, Transgenic Unit, Max Planck Institute for Molecular Genetics, Berlin, Germany. [18]U.S. Department of Energy Joint Genome Institute, Berkeley, CA 94720, USA. [19]School of Natural Sciences, University of California, Merced, CA 95343, USA. [20]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité - Universitätsmedizin Berlin, Berlin, Germany. [21]Epigenetics and Sex Development Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Berlin, Germany.
*Corresponding author. Email: mundlos@molgen.mpg.de (S.M.); dario.lupianez@mdc-berlin.de (D.G.L.)



	Contigs (PacBio)	Scaffolds (HiC)
num contigs	8,093	4,454
Largest contig	21,251,606	155,848,725
Total length	2,098,179,395	2,099,999,819
N50	2,611,711	119,847,213
GC (%)	41.91	41.91
Annotated genes		18,751
Transcripts		25,926

Fig. 1. Mole genome and epigenetic and transcriptional study of ovotestis development. (**A**) An Iberian mole (*Talpa occidentalis*) and an adult mole ovotestis. Scale bar, 500 µm. (**B**) Genome assembly of *T. occidentalis* and gene annotation statistics. (**C**) Phylogenetic tree, based on fourfold degenerate sites, between selected species. mya, million years ago. (**D**) Venn diagram of active enhancers (data S1). (**E**) Principal component (PC) analysis of RNA-seq datasets of P7 mole gonads.

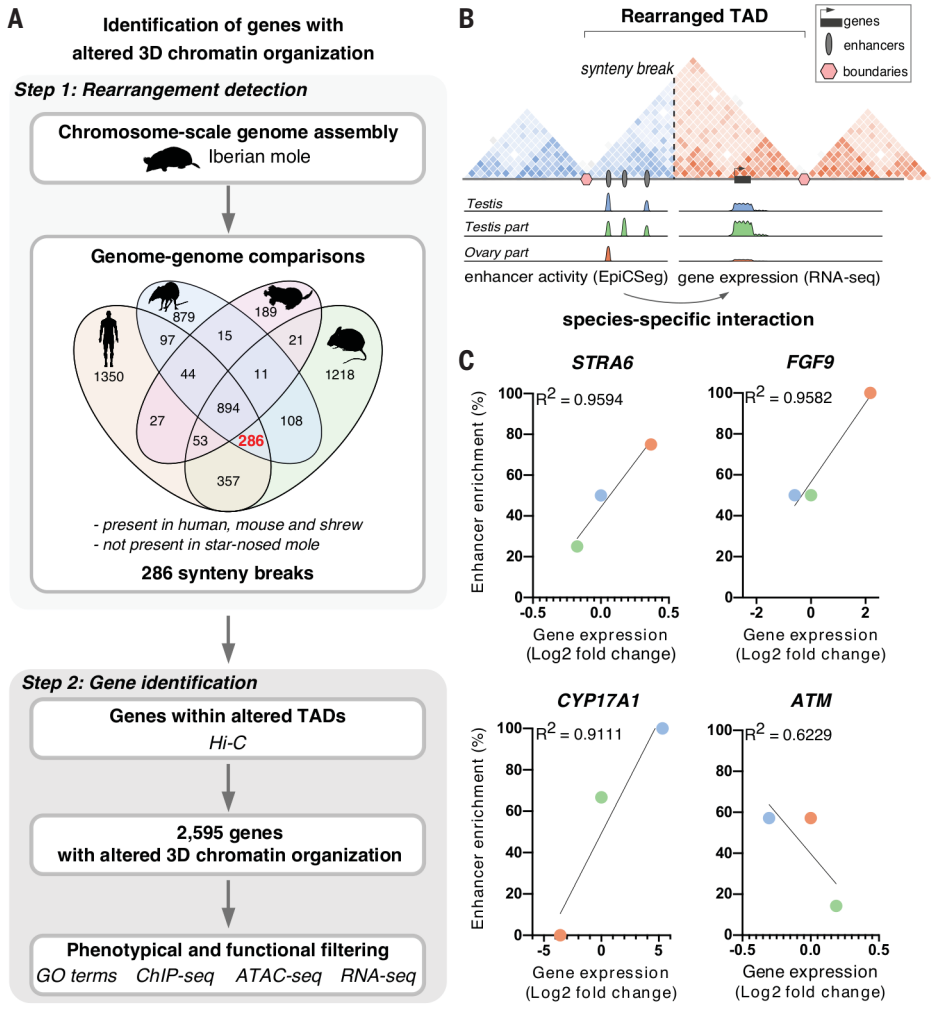


Fig. 2. Identification of genes with altered 3D chromatin regulatory landscapes. (**A**) Strategy used to identify genes with altered 3D chromatin organization as a result of species-specific rearrangements. (**B**) Strategy used to assign regulatory elements to candidate genes. Number of active enhancers is correlated to gene expression levels for each tissue. (**C**) Correlation between the percentage of active enhancers and gene expression per tissue (orange, ovary part; green, testis part; blue, male testis) for selected candidates (full gene dataset in fig. S4 and data S7). *STRA6*, *FGF9*, and *CYP17A1* display the highest positive correlation; *ATM* shows negative correlation.

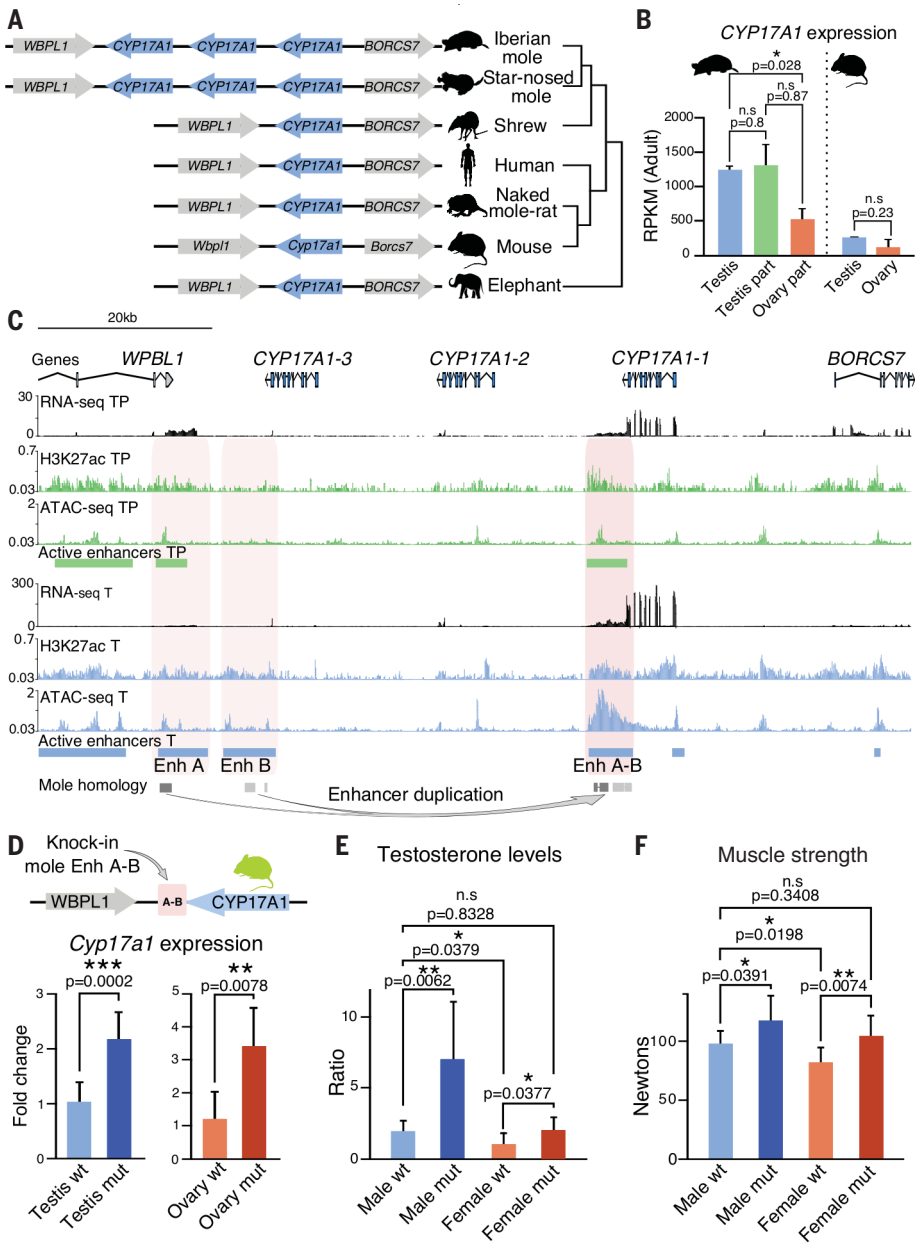


Fig. 3. Duplication of regulatory elements at the *CYP17A1* locus and associated increase in androgen production and strength. (**A**) Comparative genomics at the *CYP17A1* locus. (**B**) *CYP17A1* expression (RNA-seq) in mole and mouse adult gonads (n = 2). RPKM, reads per kilobase per million reads. (**C**) Expression profile (RNA-seq, top), enhancer marks (H3K27Ac, center), and open chromatin (ATAC-seq, bottom) for testis part and testis at P7 gonads. Segmentation for active enhancers for testis part (green bars) and testis (blue bars). BLAT Sequence homology is represented in gray boxes. Duplicated enhancer (A-B) results from fusion of enhancer A and B. (**D**) (Above) Integration of the mole *CYP17A1* duplicated enhancer (Enh) into the mouse *Cyp17a1* locus. (Below) Expression analysis of *Cyp17a1* (reverse transcription quantitative polymerase chain reaction) in adult mouse mutant gonads (mut) and wild-type controls (wt) ($n \geq 5$). (**E**) Circulating testosterone levels in adult mouse mutants and wild-type controls (n = 7). (**F**) Grip-strength test in adult mouse mutants and wild-type controls (n = 7). Bars represent mean and SD. Two-sided Student's t test. n.s, nonsignificant. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$.

enhancers (2726) indicates a distinct molecular profile from the testis and OP. Principal component analysis of RNA-seq data further confirmed the distinctive nature of the TP (Fig. 1E).

An analytical framework for evolutionary analyses

We combined our functional datasets with comparative genomics analyses and focused at three distinct levels: genes, regulatory elements, and 3D chromatin organization (fig. S4). We reasoned that phenotype-relevant mutations affecting these levels should be shared by the Iberian mole and the American star-nosed mole (*Condylura cristata*), whose females also develop ovotestes. We first searched for gene families that underwent expansion or contraction in the mole lineage, as well as genes under positive selection (supplementary text and data S2 and S3). Gene Ontology (GO) enrichment analyses revealed signatures in metabolic, immunological processes and the olfactory receptor repertoire. By filtering with GO terms related to sex differentiation ("sex, gonad") (data S4), we identified eight positively selected genes that could affect mole gonadal development (table S1). To gain functional insight into these genes, we searched for mouse- and human-reported mutations. Although some mutations led to reduced fertility, none was reported to induce sex reversal, thus making a contribution to the mole intersex phenotype unlikely (data S3).

Next, we focused on regulatory elements by identifying mole-accelerated regions, defined as genomic segments that are highly conserved during mammalian evolution but diverged in moles (supplementary text and data S5). We identified 3560 mole-accelerated regions that were subsequently filtered for overlap with our predicted gonadal enhancers (129 regions). Instead of associating the accelerated enhancers to the nearest gene, we used TAD predictions from Hi-C data to delimit a genomic range of interaction for each element and assign them to the genes in such regions. TADs are well conserved across tissues (fig. S5), consistent with previous findings (*4*). The assigned genes were further filtered with GO terms related to sex differentiation, which revealed two candidates: the transcription factor (TF) *Osr1* and the cell-cycle regulator *Cdk2* (table S1). Although both genes are essential for gonadogenesis (*16*, *17*), they show a similar expression pattern in moles and mice and are likely not relevant for mole ovotestis formation (fig. S6).

Because no relevant candidates were found, we searched for changes in 3D chromatin organization, on the basis of the hypothesis that rearrangements can alter regulatory domains and affect gene expression (*8*, *9*). Rearrangements can be identified in genome comparisons as synteny breaks, defined here as alterations on the conserved colinear order of loci between species. To identify mole-specific rearrangements, we compared the mole genomes (*T. occidentalis* and *C. cristata*) with full-chromosome assemblies from human, mouse, and shrew, as the closest taxonomical outgroup with normal ovarian development (Fig. 2A). We used our Hi-C domain predictions to identify genes located in TADs affected by a synteny break, for a total of 2595 candidate

genes considered to be susceptible to altered regulation (data S6). We filtered these candidates according to GO terms related to sex differentiation, restricting the list to 39 genes (table S1 and data S4).

Using our functional datasets, we then searched for footprints of altered gene regulation that might be the consequence of mole-specific rearrangements. We considered the nature of each rearrangement (Fig. 2B and supplementary text) and our TAD predictions, to delimit a potential region of new interactions for each candidate gene and determine the number of active regulatory elements contained within. Of the 39 candidate genes, only 17 were predicted to gain de novo interaction with regions containing active enhancers. Furthermore, we ranked the candidate genes by correlating the number of active elements with the expression levels for each tissue, as an indicative parameter of potential effects of the rearrangement on transcription (Fig. 2C, fig. S7, and data S7). A positive correlation between active enhancers and gene expression was found for 10 genes. Among the top-ranked candidates, we selected those displaying higher expression for subsequent functional validation: the androgen-related gene *CYP17A1* and the pro-testicular growth factor gene *FGF9*.

A tandem triplication at the *CYP17A1* locus is linked to increased androgen production and strength

We detected an intra-TAD tandem triplication at the mole *CYP17A1* locus that creates two additional copies of the gene. Through comparative genomics analysis, we confirmed the exclusive presence of the rearrangement in the mole lineage and its absence in other mammals (Fig. 3A and fig. S8). *CYP17A1* encodes a key enzyme controlling androgen synthesis (*18*), suggesting a role in female mole masculinization. The triplication was associated with high *CYP17A1* expression in testis and TP, both substantially exceeding the expression levels in mice (Fig. 3B and data S8). Searching for other genes of the steroidogenic pathway, we observed that *CYP19A1*, located downstream of *CYP17A1*, is not expressed in the TP (fig. S9). *CYP19A1* encodes for aromatase, an enzyme that converts androgens to estrogens (*18*), and is expressed exclusively in the OP (fig. S9). It is thus expected that the high levels of *CYP17A1* in the OP do not impede estrogen production and reproductive function, because of the protective effect of aromatase in degrading androgens locally. The absence of *CYP19A1* expression in the TP, in combination with high *CYP17A1* expression, provides a plausible explanation for the masculinization observed in female moles. We used gas chromatography–mass spectrometry to quantify serum levels of male hormones and found high levels of circulating

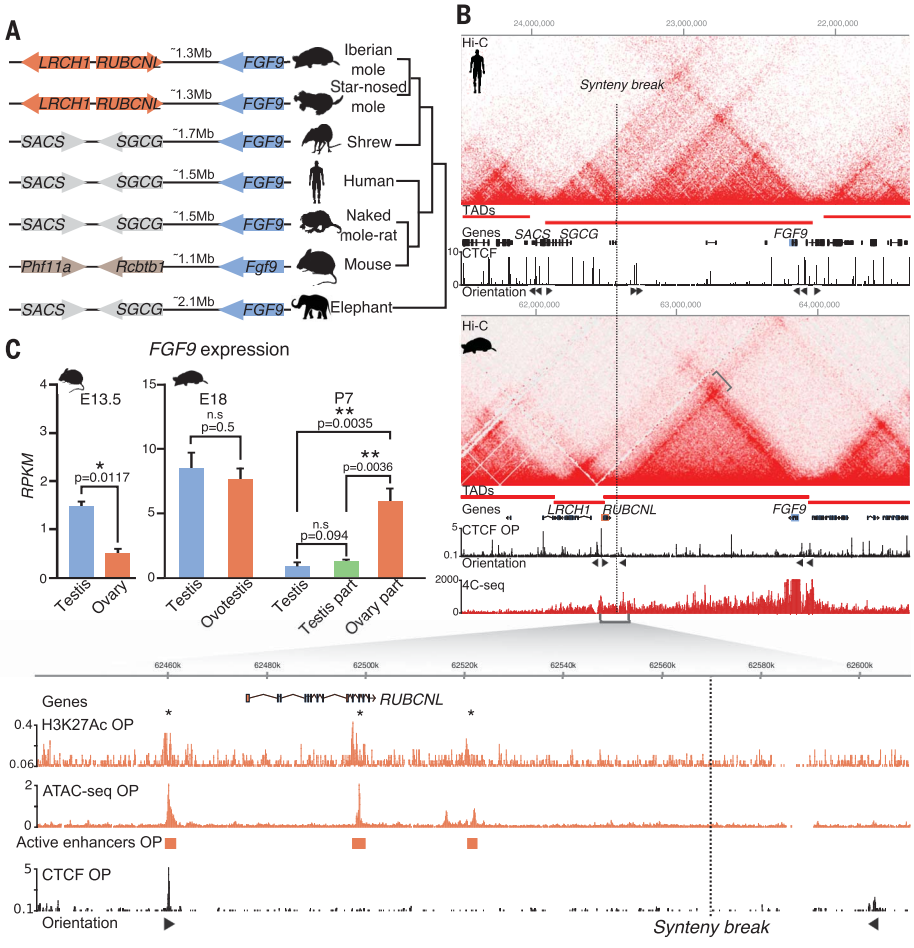


Fig. 4. An inversion altering the regulatory landscape of the *FGF9* mole locus. (**A**) Comparative genomics at the *FGF9* locus. (**B**) Hi-C maps for human and mole displaying synteny break (discontinuous line) and TAD prediction. (Below) CTCF ChIP-seq (mole P7 gonads) with peak orientation. 4C-seq using mole *FGF9* promoter as viewpoint shows contact extension beyond synteny break. Zoom of *FGF9* interacting region shows active ovarian enhancers (asterisks). H3K27Ac, ATAC-seq, and segmentation for active enhancer tracks are displayed in orange. (**C**) *FGF9* expression (RNA-seq) in mice and moles at different time points. Bars represent mean and SD ($n \geq 2$). Two-sided Student's t test. n.s = nonsignificant. $^*P \leq 0.05$, $^{**}P \leq 0.01$.

androgens in female moles. Androgen levels were similar to those in male individuals (fig. S9), contrary to the general pattern among mammals in which males display higher levels than females.

To examine the contributions of the additional *CYP17A1* alleles to increased androgen production, we analyzed RNA-seq data from mole testes and TP. The three *CYP17A1* paralogues have sufficiently diverged to enable unambiguous mapping of RNA-seq reads. The two newly emerged *CYP17A1-2* and *CYP17A1-3* jointly contribute less than 5% of the *CYP17A1* transcript (Fig. 3C). Furthermore, sequence conservation analyses revealed that they diverge more from the human sequence than *CYP17A1-1* (fig. S10). These findings suggest that the triplication of the *CYP17A1* gene itself does not explain the increased androgen levels in moles. Instead, the triplication also caused the duplication and fusion of two predicted enhancer elements, termed "enhancer A" and "enhancer B" (Fig. 3C). As the triplication does not affect any TAD boundary, the duplicated enhancers and *CYP17A1* genes locate within the same regulatory domain (fig. S11). The duplicated fusion element, "enhancer A-B," shows a high degree of sequence conservation with the original elements (84 and 90%, respectively) and high abundance of active enhancer marks (Fig. 3C). Computational predictions of binding affinities showed that enhancer A-B maintains significant binding affinities for TFs found in the original enhancer A, as well as previously uncharacterized binding affinities for TFs that are expressed in the TP (data S9). Together, these observations suggest that duplication and functional changes in regulatory sequences, rather than amplification of coding sequence, cause the observed phenotypic adaptations.

To confirm this hypothesis, we inserted the mole enhancer A-B sequence into the *Cyp17a1*

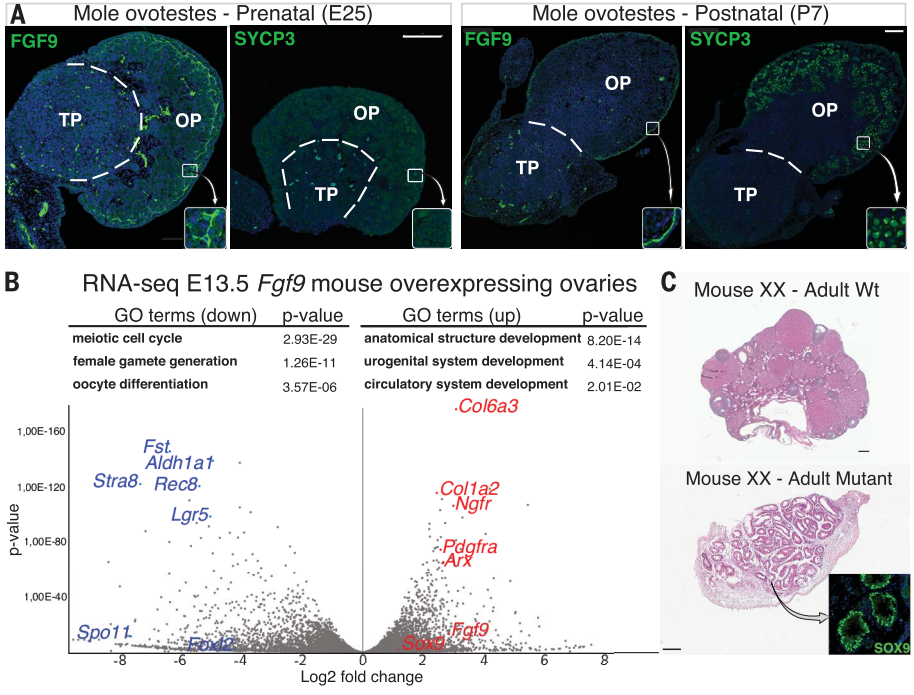


Fig. 5. *FGF9* sustained expression delays meiosis and promotes XX gonadal masculinization. (**A**) Spatiotemporal expression of FGF9 and the meiotic marker SYCP3 [immunostaining, green; 4′,6-diamidino-2-phenylindole (DAPI), blue]. Insets display zoomed regions from OP. Scale bars, 100 μm. (**B**) Volcano plot from RNA-seq of XX mutant versus wild-type gonads (E13.5) and GO analysis. (**C**) Hematoxylin and eosin staining of XX gonads of adult mutants and wild-type controls. Cord-like structures in mutants denote XX-to-XY sex reversal. Inset shows SOX9 expression (immunostaining, green; DAPI, blue). Scale bars, 200 μm.

murine locus (Fig. 3D). Adult knock-in mice carrying the additional mole enhancer displayed a three- and twofold up-regulation of *Cyp17a1* expression in females and males, respectively. The increased expression occurs in the same cell type (steroidogenic cells) as in wild-type controls (fig. S12). A similar effect was observed at embryonic stages (fig. S12), thus confirming the increased in vivo activity of the mole-specific fusion enhancer in gonadal tissue. The up-regulation of *Cyp17a1* expression was accompanied by an increase in circulating testosterone in females and males (two- and threefold respectively) (Fig. 3E). Because androgens have potent anabolic effects in muscle, we performed a grip-strength test that revealed a significant increase in muscle strength in mutants compared with wild-type controls (Fig. 3F). Thus, the regulatory nature of the *CYP17A1* rearrangement offers a plausible molecular mechanism for the observed shift in hormone levels and the corresponding phenotype.

An inversion at the *FGF9* locus is associated with delayed meiosis and gonadal masculinization

We identified a rearrangement at the *FGF9* locus, which is exclusive to the mole lineage and not present in any other mammals examined (Fig. 4A and figs. S13 and S14). *FGF9* is a known testis-determining gene that functions in positive feedback with *SOX9* and inhibits the ovary-determining *WNT4*/β-catenin pathway (*11*) (fig. S15). Consequently, loss of *Fgf9* in XY gonads results in down-regulation of testicular markers and male-to-female sex reversal. Comparative analyses against human, representative of the ancestral organization at the locus, revealed a large inversion that relocates a distant genomic region (26 Mb away in the human genome) to the mole *FGF9* locus (Fig. 4B). Hi-C data showed that the synteny break occurs in the human *FGF9* TAD, disrupting its 3D organization. Thus, the mole locus is reorganized, with most of the *FGF9* TAD remaining conserved but extending beyond the synteny breakpoint on the centromeric side. This extended interaction domain is delimited by the presence of two CTCF binding sites with divergent orientation, a genomic signature associated with TAD boundaries (*7*) (Fig. 4B). A closer examination of the newly interacting region revealed several elements enriched for active epigenetic marks, some of them specific for the mole ovotestis (fig. S16). This interaction pattern was also validated through independent circular chromosome conformation capture sequencing (4C-seq) experiments (Fig. 4B), which confirmed the association of the *FGF9* gene with the region containing the putative tissue-specific enhancers. We tested one element (fig. S16) in mouse LacZ reporter assays. Although this element displayed enhancer activity in tissues such as the eye (fig. S17), we did not observe gonadal staining, which could be due to specific requirements for additional trans-acting factors in the mole. Alternatively, it may reflect known limitations of the reporter assay that might be intensified by the interspecies nature of the experiment (*19*).

We then explored possible alterations on the dimorphic *FGF9* expression pattern observed in other mammals, which is essential for suppressing germ cell meiosis (*20*). In mice, *Fgf9* is first expressed in the bipotential gonad of both sexes and becomes progressively restricted to the testis and turned off in ovaries, allowing the initiation of female meiosis at embryonic day 13.5 (E13.5). In mole gonads, however, *FGF9* expression is maintained after sex determination in both sexes and becomes restricted to the OP at later stages (Fig. 4C). Immunostaining analyses showed that FGF9 expression persists in the OP across the entire prenatal period and becomes confined to a thin rim at postnatal stages (P7). The spatial reduction in FGF9 expression is concomitant with the initiation of meiosis (Fig. 5A and fig. S18), which is considerably delayed in female moles until birth, an exceptional feature among mammals (*21*). Consequently, the observed heterochrony on mole *FGF9* expression, compared with mouse, and its potential effects are suggestive of a contribution to the masculinization of female mole gonads.

We hypothesized that *FGF9* expression during early XX mole gonadogenesis might prevent germ cells from entering meiosis in the OP, allowing the TP to develop further. To test this hypothesis, we engineered a bacterial artificial chromosome (BAC) construct to overexpress *Fgf9* in somatic ovarian populations and generated transgenic mice through *PiggyBac* transgenesis and morula aggregation. Highly chimeric animals displayed early embryonic lethality, likely due to *Fgf9* misexpression in other organs, an effect that precluded their study in later stages. Nevertheless, RNA-seq analysis of E13.5 ovaries demonstrated an inhibition of the meiotic process, manifested by down-regulation of meiosis markers (Fig. 5B and data S10). By contrast, low-chimera individuals composed of XX wild-type and XX mutant cells were viable and displayed *Fgf9* expression during the entire ovarian development. These animals showed a complete female-to-male sex reversal, defined by gonadal morphology and expression of male-specific factors such as SOX9 (Fig. 5C). These results directly confirm the potential of altered *FGF9* expression to induce masculinization in mammalian XX females.

Discussion

Vertebrate sex determination is controlled by a limited set of key regulators whose hierarchy has evolved dynamically (*11*). Most of these genes display pleiotropic effects, controlling regulatory networks in several tissues, often making them indispensable for embryonic viability (*22*, *23*). It is thus plausible that variations in sex determination derive from regulatory changes that alter gene expression patterns but preserve essential functions, as suggested for other evolutionary adaptations (*1*, *2*). These genomic changes appear to be linked to the evolutionary success of moles and demonstrate that regulatory innovation can overcome a priori seemingly incompatible situations such as female fertility in the presence of high androgen levels. But how could an intersex phenotype evolve, and what is its adaptive significance? One potential explanation is the anabolic effects of androgens on muscle mass. Mole ovotestes may have evolved to equalize muscular strength among sexes by increasing androgens in females. Such an advantage, combined with other androgen-derived effects such as aggressive behavior, might have been key for the adaptation to the demanding requirements of a burrowing underground lifestyle (*14*).

This study also highlights the evolutionary importance of genomic rearrangements and their potential to modulate developmental gene expression. In most cases, genomic rearrangements would have limited effects, as they preserve whole regulatory units and do not disrupt enhancer-promoter interactions (*6*, *24*). However, as shown here, they can also alter the regulatory potential in the local environment of a TAD, as for the *CYP17A1* locus, or shuffle the functional content of distant TAD units, exemplified by the *FGF9* locus. Similar effects have been observed in human genetic diseases for enhancer duplications (*25*) and for inversions (*8*, *9*). Therefore, regulation of genes by enhancer elements and their organization in TADs might constitute a mechanism of "evolvability" resulting from a modular system with vast flexibility and enormous evolutionary potential for the origin of novelties. On the basis of this modularity, genomic rearrangements can easily change and reconstitute complex expression patterns, thereby contributing to the saltatory nature of phenotypic innovation observed in many phylogenetic lineages. We expect that approaches considering these important aspects will eventually reveal the evolutionary basis of many other traits and substantially enhance the toolbox for unlocking the secrets of phenotypic variation and adaptation across the animal kingdom.

REFERENCES AND NOTES

1. J. Lopez-Rios *et al.*, *Nature* **511**, 46–51 (2014).
2. Y. F. Chan *et al.*, *Science* **327**, 302–305 (2010).
3. O. Dudchenko *et al.*, *Science* **356**, 92–95 (2017).
4. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
5. E. P. Nora *et al.*, *Nature* **485**, 381–385 (2012).
6. N. Harmston *et al.*, *Nat. Commun.* **8**, 441 (2017).
7. C. Gómez-Marín *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7542–7547 (2015).
8. M. Spielmann, D. G. Lupiáñez, S. Mundlos, *Nat. Rev. Genet.* **19**, 453–467 (2018).
9. D. G. Lupiáñez *et al.*, *Cell* **161**, 1012–1025 (2015).
10. M. Franke *et al.*, *Nature* **538**, 265–269 (2016).
11. B. Capel, *Nat. Rev. Genet.* **18**, 675–689 (2017).
12. F. J. Barrionuevo, F. Zurita, M. Burgos, R. Jiménez, *Dev. Biol.* **268**, 39–52 (2004).
13. F. D. Carmona *et al.*, *J. Exp. Zool.* **310B**, 259–266 (2008).
14. J. Haeck, *Neth. J. Zool.* **19**, 145–248 (1969).
15. C. J. Douady *et al.*, *Mol. Phylogenet. Evol.* **25**, 200–209 (2002).
16. Q. Wang, Y. Lan, E.-S. Cho, K. M. Maltby, R. Jiang, *Dev. Biol.* **288**, 582–594 (2005).
17. C. Berthet, E. Aleem, V. Coppola, L. Tessarollo, P. Kaldis, *Curr. Biol.* **13**, 1775–1785 (2003).
18. I. Hanukoglu, *J. Steroid Biochem. Mol. Biol.* **43**, 779–804 (1992).
19. K. Suryamohan, M. S. Halfon, *WIREs Dev. Biol.* **4**, 59–84 (2015).
20. J. Bowles *et al.*, *Dev. Cell* **19**, 440–449 (2010).
21. F. Zurita *et al.*, *Sex Dev.* **1**, 66–76 (2007).
22. J. S. Colvin, A. C. White, S. J. Pratt, D. M. Ornitz, *Development* **128**, 2095–2106 (2001).
23. W. Bi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6698–6703 (2001).
24. N. H. Lazar *et al.*, *Genome Res.* **28**, 983–997 (2018).
25. A. J. Will *et al.*, *Nat. Genet.* **49**, 1539–1545 (2017).

ACKNOWLEDGMENTS

We thank A. Stiege, N. Brieske, U. Fisher, K. Macura, J. Fiedler, K. Zill, C. Franke, and G. Mastrobuoni for technical support. We thank P. Koopman for providing us with the Wt1-BAC vector. We thank the sequencing and genotyping facility of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden for helping with the initial long-read sequencing project. We thank the Animal Outcome Core Facility of the NeuroCure Center, Charité University, Berlin, for helping with strength tests. We also thank members of the Lupiáñez and Mundlos labs for fruitful discussions. **Funding:** This research was supported by grants from the Deutsche Forschungsgemeinschaft (grant numbers MU 880/15-1 and MU 880/16-1) and Max Planck Society. D.G.L. was supported by the Fundación Alfonso Martín Escudero and by a Helmholtz ERC Recognition Award grant from the Helmholtz-Gemeinschaft (ERC-RA-0033). Work conducted at the E. O. Lawrence Berkeley National Laboratory was performed under Department of Energy contract DE-AC02-05CH11231, University of California. A.V. and M.O. were supported by NIH grant R01HG003988. M.O. was supported by Swiss National Science Foundation grant PCEFP3_186993. O.S. was supported by the Austrian Science Fund FWF grant P32190. **Author contributions:** F.M.R., S.M., and D.G.L. designed the experimental approach. F.M.R., D.G.L., F.B., and R.J. captured moles and prepared the samples. F.M.R., M.S., and D.G.L. processed the samples for DNA, Hi-C, RNA, ChIP, and ATAC sequencing. S.A.H., B.T., A.M., and M.V. supervised bioinformatic analyses. R.S., D.H., M.-H.M., and H.K. assembled the genome. P.F., P.X., and O.S. performed the gene annotation and posterior genomic analyses. S.A.H. and A.B. analyzed the RNA-seq data. S.A.H. performed the epigenetic analysis and the prediction of synteny breaks. M.H. performed multispecies alignments. R.S., V.H., and T.K. analyzed chromatin interaction datasets. M.F.H. and S.A.W. performed hormone measurements. F.M.R., D.G.L., and A.H. performed the *CYP17A1* enhancer transgenic experiment. F.M.R., I.H., and M.O. performed the *FGF9* transgenic experiments. L.W. derived the XX mESC line and performed blastocyst aggregations. F.M.R., M.S., S.M., and D.G.L. performed ELISA experiments and grip-strength tests. F.M.R., D.K.N.D., A.V., A.M., S.M., and D.G.L. participated in data interpretation and discussion. F.M.R., S.M., and D.G.L. wrote the manuscript with contributions from all other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number PRJNA494291. All transcriptomes and epigenetic data have been deposited in GEO under accession number GSE120589. Alignments have been deposited at https://bds.mpi-cbg.de/hillerlab/IberianMole.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/208/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S8
References (*26*–*100*)
MDAR Reproducibility Checklist
Data S1 to S10

View/request a protocol for this paper from *Bio-protocol*.

27 August 2019; resubmitted 19 April 2020
Accepted 17 August 2020
10.1126/science.aaz2582

REPORT

ORGANIC CHEMISTRY

Electro-inductive effect: Electrodes as functional groups with tunable electronic properties

Joon Heo[1,2]*, Hojin Ahn[1]*, Joonghee Won[1,2]*, Jin Gyeong Son[3], Hyun Kyong Shon[3], Tae Geol Lee[3], Sang Woo Han[1]†, Mu-Hyun Baik[2,1]†

In place of functional groups that impose different inductive effects, we immobilize molecules carrying thiol groups on a gold electrode. By applying different voltages, the properties of the immobilized molecules can be tuned. The base-catalyzed saponification of benzoic esters is fully inhibited by applying a mildly negative voltage of −0.25 volt versus open circuit potential. Furthermore, the rate of a Suzuki-Miyaura cross-coupling reaction can be changed by applying a voltage when the arylhalide substrate is immobilized on a gold electrode. Finally, a two-step carboxylic acid amidation is shown to benefit from a switch in applied voltage between addition of a carbodiimide coupling reagent and introduction of the amine.

In 1937, Hammett reported his groundbreaking studies on how functional groups affect the rate of chemical reactions (*1*). This work led to the broadly accepted and widely used concept of the inductive effect of functional groups to tune the electronic properties of molecules. Of the countless applications of this concept in chemistry, the ability to control the rates of stoichiometric and catalytic reactions by altering the functional groups that are covalently bound to reactants or catalysts is perhaps the most prominent and pervasive use of inductive effects. Despite being a cornerstone of chemistry at large, this approach to modulating the property of a molecule is neither efficient nor convenient. Each functional group can only impose an inductive effect that is specific to its identity, necessarily creating a discontinuous series where the range of properties is limited by what the functional groups offer. Naturally, each functionalized derivative must be prepared, which is laborious and limiting, because not all possible molecular compositions are easily accessible.

We envisioned a different approach for modulating the electronic properties of a molecule, whereby instead of preparing a series of derivatives, the molecule of interest may be covalently attached to an electrode and its chemical properties modulated by changing the applied voltage (Fig. 1). Applying a negative potential should mimic electron-donating groups, whereas a positive potential should reproduce the electronic effect of electron-withdrawing groups. In this paradigm, the inductive effect is no longer discrete, because the applied voltage can be adjusted continuously. Recently, Dawlaty and others showed in a thorough spectroscopic study that such a correlation between Hammett parameters and polarization of the molecule exists (*2*). Specifically, the vibrational frequency shifts of para-substituted mercaptobenzonitrile immobilized on a gold electrode could be directly correlated to the applied voltage. Although this study did not explore control of chemical reactions using voltage, the connection between applied voltage and Hammett parameters directly supports our proposed strategy. If the electrode could act as a functional group with adjustable inductive effect, it would no longer be necessary to prepare a series of analogous molecules or catalysts to impose different inductive effects on the reactive core. A particularly attractive feature of using an electrode in this way is that the inductive effect applied to the molecule can be varied over time, thereby modulating the electronic structure of a catalyst over the course of a catalytic cycle, for example. This feature could enable fine control over the progress of chemical reactions by application of voltage sequences, reminiscent of pulse sequences in nuclear magnetic resonance spectroscopy.

The idea of covalently attaching molecules to an electrode is, of course, not new, but immobilization efforts thus far have been aimed at reactions where the deposited molecules participate in redox reactions. Molecules are typically immobilized on an electrode because their reactivity hinges on receiving or donating electrons. The key feature of our approach, which we term the electro-inductive effect, is that an electrode can modulate nonredox chemical reactivity. That the electric field induced by an applied voltage can alter the electronic structure of immobilized molecules on the electrode is known as the Stark effect (*3*, *4*), but the electrode was not used as a functional group to control chemical reactions of reactants that were immobilized on the surface.

To prove the principle, we explored whether the electro-inductive effect could change the rate of the base-mediated saponification of a benzoic ester. This reaction served as one of the 39 standard reactions from which the Hammett parameters were originally derived (*1*). As highlighted in Fig. 2A, a benzoic ester carrying a thiol moiety at the para position could readily be self-assembled as a monolayer on a gold surface (*5*). Nucleophilic addition of a hydroxide ion to the carbonyl of the ester affords the hydrolyzed products: the benzoate anion and the corresponding alcohol. It is easy to understand that electron-withdrawing groups on the substrate make a carbonyl carbon more electrophilic, leading to an acceleration of the saponification reaction, whereas electron-donating groups diminish the reaction rate. Our initial studies showed that the saponification of the immobilized benzoates containing methyl and ethyl esters was problematic, which can be attributed to too tightly packed monolayers that are protected against chemical attacks (*6*, *7*). Thus, we used a sterically bulkier *tert*-butyl ester, *tert*-butyl 4-mercaptobenzoate (TBMB), to form a slightly less dense monolayer, allowing the esters to undergo chemical reactions more easily. The progress of saponification was conveniently monitored by surface-enhanced Raman spectroscopy (SERS) techniques (*8*).

If the electro-inductive effect were substantial, the saponification of TBMB to generate 4-mercaptobenzoate should be accelerated when a positive voltage is applied, whereas the reaction should become slower with a negative voltage, as illustrated in Fig. 2A. The SERS spectra of the reactant and the product show two common peaks assigned to the breathing modes of the aromatic ring [ν(C–C)$_{ring}$] with very strong intensities at ~1070 (*9*) and ~1580 cm^{-1} (Fig. 2B). In addition, other distinct peaks at 1383, 1706, and 1405 cm^{-1} correspond to C(CH$_3$)$_3$, C=O stretching mode [ν(C=O)] in the reactant, and COO$^-$ stretching mode [ν_s(COO$^-$)] in the product, respectively (*10*–*12*). As the reaction proceeds, we expect a decrease in peak intensities of the vibrations related to C(CH$_3$)$_3$ and C=O and an increase in peak intensity for ν_s(COO$^-$). However, because the SERS peaks for C(CH$_3$)$_3$ and ν_s(COO$^-$) appear in a similar region around 1400 cm^{-1}, it proved difficult to precisely monitor the reaction progress using these peaks. Therefore, the moderately strong

[1]Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea. [2]Center for Catalytic Hydrocarbon Functionalizations, Institute for Basic Science (IBS), Daejeon 34141, Republic of Korea. [3]Center for Nano-Bio Measurement, Korea Research Institute of Standards and Science (KRISS), Daejeon 34113, Republic of Korea.
*These authors contributed equally to this work.
†Corresponding author. Email: mbaik2805@kaist.ac.kr (M.-H.B.); sangwoohan@kaist.ac.kr (S.W.H.)

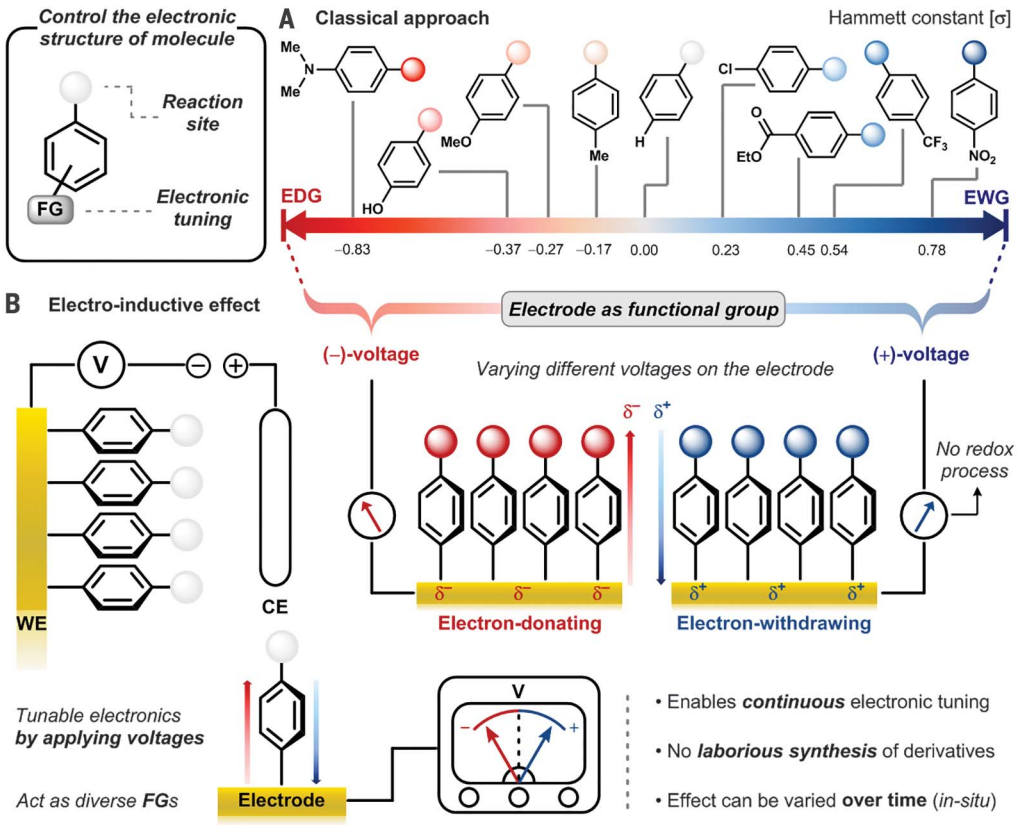


Fig. 1. Proposed design for exploiting the electro-inductive effect to control chemical reactions. (**A**) Classical approach for modulating the electronic properties of a molecule. FG, functional group; EDG, electron-donating group; EWG, electron-withdrawing group; Me, methyl; Et, ethyl. (**B**) Electronic tuning of the reaction site with covalent attachment of a molecule to an electrode at an applied potential. WE, working electrode; CE, counter electrode; δ^- and δ^+, partial negative charge and partial positive charge.

peak at 1706 cm^{-1}, which is instantly recognizable as a C=O vibration in the reactant, was chosen as an indicator for monitoring the reaction progress. The intensities of the SERS signals are difficult to use as a quantitative measure of concentration in an absolute sense without careful calibration; the evolution of these signals over the reaction time of the current experiment relative to an internal standard is meaningful and serves as a qualitative indicator for the reaction progress (*13*, *14*).

Before applying an external voltage, an appropriate potential window was ascertained through cyclic voltammetry to confirm the absence of electron transfer events (see figs. S3 to S5). Specifically, potentials of +0.25 and −0.25 V versus open circuit potential (OCP) were used. Figure 2C shows the SERS spectra of TBMB immobilized on the gold electrode before and after 22.5 hours at 50°C in 3 M potassium hydroxide solution with no voltage, +0.25 V, or −0.25 V applied. By comparing the vibrational signature of the ester at 1706 cm^{-1} to the prominent peak at ~1070 cm^{-1} [$\upsilon(C–C)_{ring}(1)$] as an internal reference, the ratios of the peak intensities could be determined, as illustrated in Fig. 2D.

Before the surface-bound substrates were exposed to the saponification conditions, the ester-associated peak was clearly visible, and the peak intensity ratio relative to the peak of $\upsilon(C–C)_{ring}(1)$ was found to be 0.183. At OCP—that is, when no voltage was applied—the peak at 1706 cm^{-1} diminished substantially after exposure of the surface-bound esters to the hydrolysis conditions, as shown in black in Fig. 2C, and the peak ratio was found to be 0.072, indicating a considerable depletion of esters on the surface. In good agreement with the aforementioned expectation, the reaction could be accelerated by applying a positive voltage. At +0.25 V, the ester peak at 1706 cm^{-1} dramatically reduced over the reaction time of 22.5 hours, as shown in blue in Fig. 2C. By contrast, the application of a negative voltage of −0.25 V led to a nearly complete inhibition of the reaction, indicated by the SERS spectrum shown in red in Fig. 2C that is practically identical to the spectrum of the initial state before exposure to the saponification conditions, which is displayed in green. The relative peak intensity ratio was found to be 0.181 in this case, very close to the ratio of 0.183 determined for the sample before hydrolysis, indicating that most of the esters were still intact on the surface. These observations suggest that the degree of saponification can be controlled by the applied potential, as proposed. The complete inhibition of the ester hydrolysis using what may be considered a mildly negative voltage of −0.25 V is notable, given that the reaction conditions, including a concentrated 3 M solution of potassium hydroxide, are highly favorable for saponification.

There is an alternative explanation for these experimental observations to consider. When a negative potential is applied to the gold electrode, all anionic species, including the hydroxide reactant, are repelled from the surface, whereas cationic species are drawn close to the surface. As a result, the local concentration of anions drops substantially, whereas the concentration of cations increases, establishing the electrochemical double layer. Thus, the lack of saponification might simply be a result of diminished hydroxide concentration within the double layer. A reasonable way of estimating the width of the double layer is to use the Gouy-Chapman model (*15*). Under the applied conditions of −0.25 V and a KOH concentration of 3 M, the width of this double layer can be estimated to be in the range of ~2 Å. With the sulfide moiety in direct contact with the gold surface, the reactive ester functionality of the TBMB substrate is calculated to be positioned 7 to 9 Å away from the surface, which is well within the diffusion layer where ionic species, such as the hydroxide ions, should be able to approach the ester group without being affected by the applied potential. Details of these calculations are given in the supplementary materials. Moreover, the examples discussed below are not affected

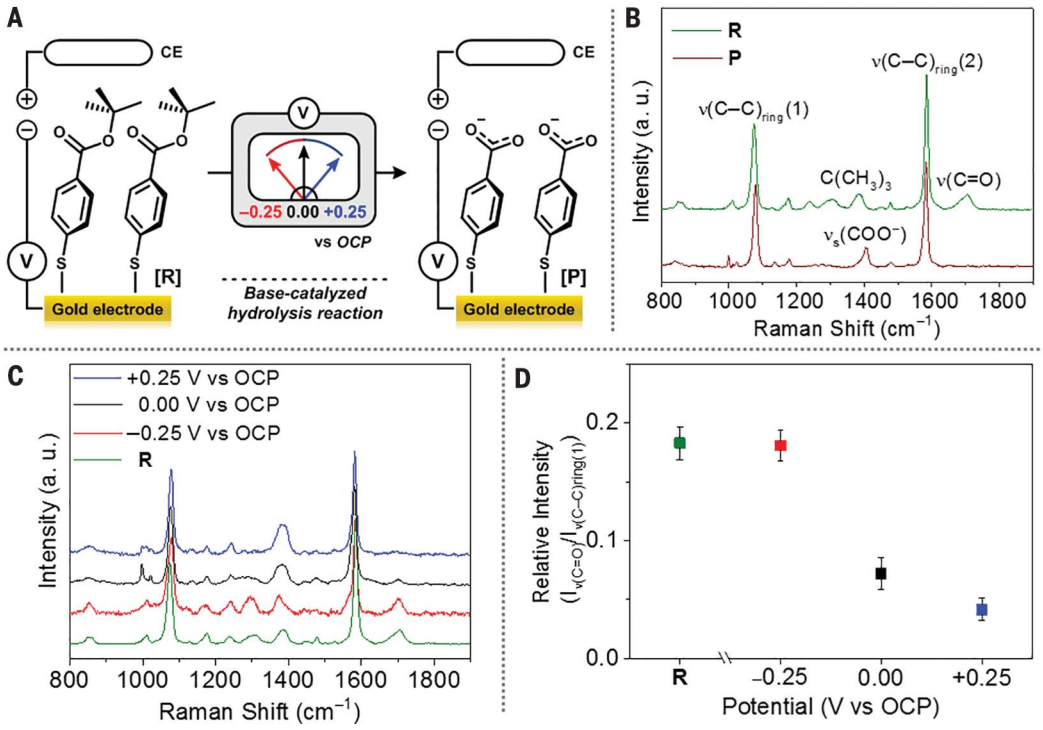


Fig. 2. Base-catalyzed ester hydrolysis reaction. (**A**) Schematic representation for base-catalyzed saponification of self-assembled monolayer of TBMB under applied potentials. Reactions were run in 3 M $KOH_{(aq)}$ at 50°C for 22.5 hours. (**B**) SERS spectra of the reactant (R) TBMB (green line) and the product (P) 4-mercaptobenzoate (brown line, deprotonated form). a.u., arbitrary units; ν, stretching mode; ν_s, symmetric stretching mode. (**C**) Representative SERS spectra of base-catalyzed saponification with +0.25 V versus OCP (blue line), OCP (black line), −0.25 V versus OCP (red line), and R (green line). (**D**) Relative intensities of base-catalyzed saponification under different potentials. The points and error bars in the relative intensity graph (D) indicate the means and standard deviations of five to seven experiments, respectively. I, intensity.

by the double-layer effect, so this alternative explanation can be excluded with some confidence.

With the initial proof-of-principle study in hand, we sought to test the electro-inductive effect in a catalytic reaction. There are two possible approaches for carrying out catalytic reactions while taking advantage of the electro-inductive effect: Either the catalyst or the reactant could be immobilized on the gold electrode and their activity changed by applying different voltages. Immobilizing the catalyst poses several practical challenges such as low catalyst loading, owing to the relatively small size of the gold electrode surface or the synthetic difficulties of installing a thiol linkage on the catalyst. Immobilizing the reactant and monitoring its transformation to the product on the gold surface is more convenient. As an illustrative example, we chose the Pd-catalyzed Suzuki-Miyaura cross-coupling reaction (*16*), which is one of the most powerful and widely used synthetic tools for C–C bond formation. Mechanistically, it is thought to proceed through three key steps: (i) oxidative addition of an aryl halide to give a Pd(II)-intermediate, (ii) transmetalation of the arylboron species that delivers the second aryl moiety to the Pd center, and (iii) reductive elimination to afford a C–C coupled biaryl product and the Pd(0) catalyst.

The SERS spectrum of the immobilized 4-bromobenzenethiol shown in Fig. 3B displays three clearly distinguishable features (*17*): (i) a C–Br stretching mode ν(C–Br) at 1068 cm^{-1} (*18*), (ii) aromatic ring vibrations $\nu(C–C)_{ring}(1)$ at 1079 cm^{-1}, and (iii) another aromatic ring vibrational mode $\nu(C–C)_{ring}(2)$ at 1560 cm^{-1} (*13, 19*). We also immobilized the anticipated product biphenyl-4-thiol on the gold electrode to confirm its SERS signature. Characteristic $\nu(C–C)_{ring}$ peaks of the biphenyl product that were most convenient to monitor could be found at 1586 cm^{-1} [$\nu(C–C)_{ring}(3)$] and 1600 cm^{-1} [$\nu(C–C)_{ring}(4)$], as illustrated in Fig. 3B (*20*). The SERS spectra of the starting material and the spectra taken after 12 hours under the Suzuki-Miyaura conditions applying +0.30 V, +0.15 V, no voltage, and −0.15 V are overlaid in Fig. 3C. After 12 hours of reaction time, the product peak of $\nu(C–C)_{ring}(3)$ can be identified very clearly, whereas the peak of $\nu(C–C)_{ring}(4)$ has become a shoulder feature. The cross-coupling reaction was slowest at +0.30 V, indicated by the weakest product peak among the series, but the rate increased notably over the series of +0.15 V, OCP, and −0.15 V, suggesting that negative induction increases the rate of the reaction. As stated above, the SERS signals cannot be used in a quantitative sense to precisely determine the reaction progress, but by comparing the two aromatic ring vibrations [$\nu(C–C)_{ring}(1)$ and $\nu(C–C)_{ring}(3)$], the trend that negative potential accelerates the reaction can be observed at least semiquantitatively. Moreover, the consumption of reactants on the surface leads to a decrease of the ν(C–Br) peak against the same internal reference, as illustrated in Fig. 3D. To complement the SERS analysis, we used time-of-flight secondary ion mass spectrometry (TOF-SIMS) techniques and analyzed the amounts of the starting material and the product on the surface. As shown in fig. S7, the peaks of the reactant ($[C_6H_4S^{79}Br]^-$) and the product ($[C_{12}H_9S]^-$) are easily distinguished in the mass spectrum. In excellent agreement with the SERS results, the peak intensity ratio of the product to the starting material increases as the applied potential is set more negative (fig. S8).

Finally, we monitored the SERS spectra of the reaction over a time course of 3 hours, taking measurements in 30-min intervals, as illustrated in Fig. 3E. As expected, the product peak increased most rapidly at −0.15 V, reaching a saturation plateau after ~90 min, whereas at other voltages, the reaction was not complete after 3 hours. Taken together, these results demonstrate that the rate of the Suzuki-Miyaura cross-coupling changes notably when the applied voltage is varied. More specifically, applying a negative voltage, to mimic an electron-donating group, enhances the rate of the reaction. The mechanism of the Suzuki-Miyaura reaction is conceptually well understood, but many details remain unclear. Indeed, the detailed mechanism of cross-coupling reactions depends on the reaction conditions, for instance, the concentration of substrate or coupling partner, the catalyst, and the base (*21–23*). Our results are in good agreement with a conventional Hammett study that used para-substituted arylbromide substrates (*24*), which showed the same trend of an increase in the

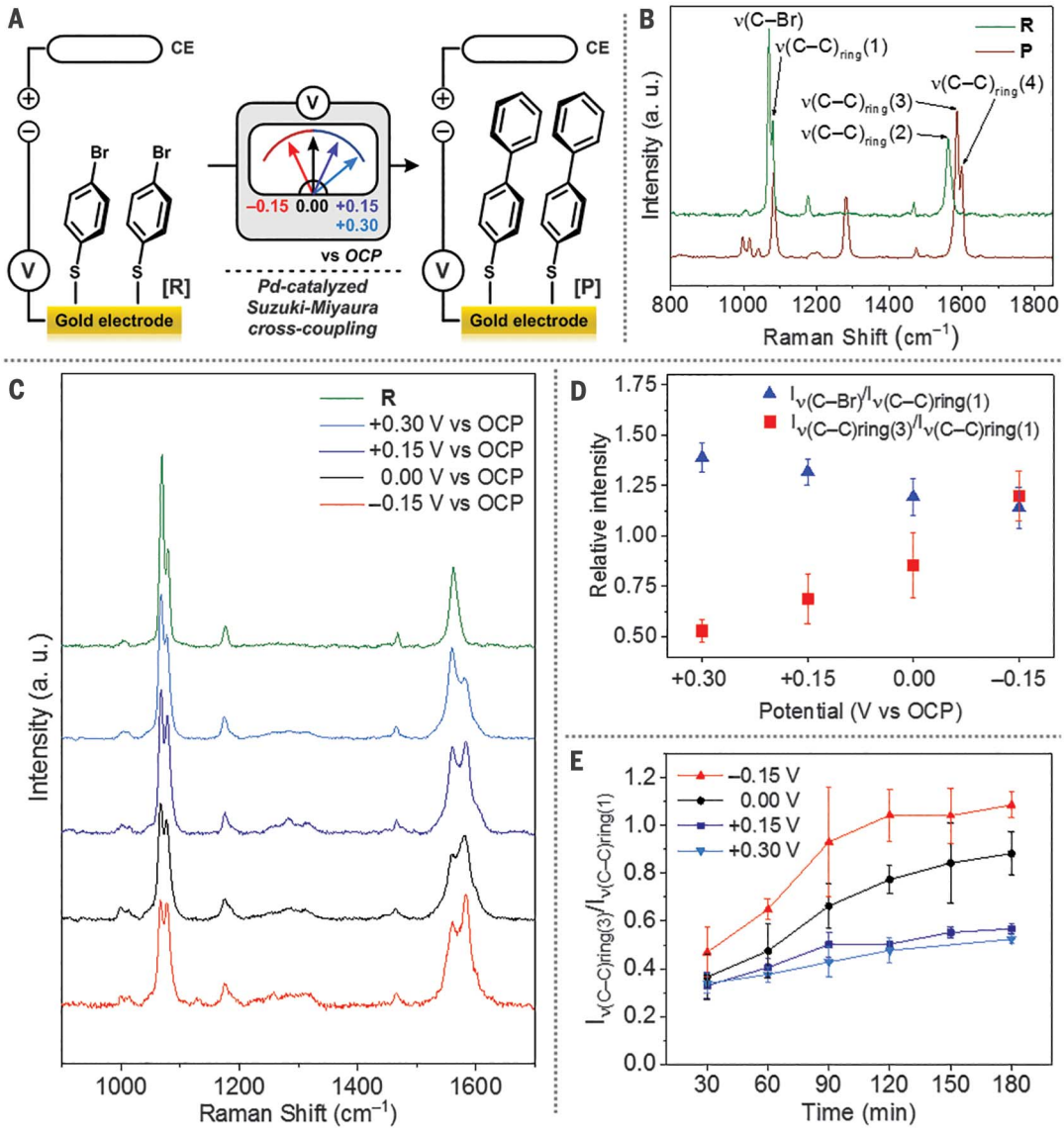


Fig. 3. Palladium-catalyzed Suzuki-Miyaura cross-coupling reaction. (**A**) Schematic representation for Pd-catalyzed Suzuki-Miyaura cross-coupling of self-assembled monolayer of 4-bromobenzenethiol under applied potentials. Reactions were run with $PhB(OH)_2$ (1.5 mmol), $Pd(OAc)_2$ (0.075 mmol), (2-biphenyl)di-*tert*-butylphosphine (0.15 mmol), KF (4.5 mmol), and tetrabutylammonium hexafluorophosphate (0.1 M) in tetrahydrofuran (3 ml) at 25°C for 12 hours (*26*). (**B**) SERS spectra of the reactant 4-bromobenzenethiol (green line, R) and product biphenyl-4-thiol (brown line, P). (**C**) Representative SERS spectra of Suzuki-Miyaura cross-coupling with R (green line), +0.30 V versus OCP (sky blue line), +0.15 V versus OCP (royal blue line), OCP (black line), and −0.15 V versus OCP (red line). (**D**) Relative intensities of Suzuki-Miyaura cross-coupling under different potentials. (**E**) Relative intensities of P in a time course under different potentials. The points and error bars in the relative intensity graphs [(D) and (E)] indicate the means and standard deviations of three to five experiments, respectively.

rate with electron-donating groups. Whereas the mechanistic interpretation of these results for the Suzuki-Miyaura reaction requires additional work, we establish that catalytic reaction rates can be modulated by the electro-inductive effect.

Encouraged by these two successful demonstrations of the electro-inductive effect, we sought to test the idea of first enhancing the nucleophilicity of the immobilized molecule and then increasing electrophilicity by inverting the applied voltage. For this purpose, an amidation mediated by a coupling reagent was chosen (*25*). As outlined in Fig. 4, a carboxylic acid such as benzoic acid first attacks 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) at the carbodiimide carbon to give an O-acylurea. This first step should be accelerated if the oxygen nucleophilicity of the carboxylic acid is enhanced. Next, the carbonyl carbon in the O-acylurea may act as an electrophile to react with an amine to afford the desired amidated product. In this second step, higher electrophilicity of the carbon should increase the reaction rate. Thus, the EDC-mediated amidation should be facilitated by first applying a negative potential and then switching to a positive potential, as conceptualized in Fig. 4A.

The SERS spectrum of the reactant 4-mercaptobenzoic acid is shown in Fig. 4B and reveals representative peaks for $\nu_s(COO^-)$ and ν(C=O) at ~1400 and ~1700 cm^{-1}, respectively. First, we confirmed that under the given conditions with all reagents present in solution, no reaction occurs if no voltage is applied, indicated by a stable and unchanged SERS spectrum of the reactant 4-mercaptobenzoic acid, as shown in Fig. 4B. Similarly, applying a voltage of +0.40 V for 1 hour had no effect, which is expected because making the benzoic acid less electron-rich should inhibit the urea formation. Applying a negative potential gives rise to a new peak at 1772 cm^{-1}, which is attributed to the carbonyl stretching mode of the O-acylurea, as is discussed in greater detail in the supplementary materials. The SERS spectrum acquired after applying a voltage of −0.70 V for 1 hour clearly shows this new carbonyl peak (Fig. 4C). Applying the −0.70 V for only 40 min and acquiring the spectrum after 20 min of applying no voltage gives a strong signal as well, and we can see a second feature at 2228 cm^{-1} as a tiny peak. This is a characteristic signature of a C≡N stretching mode [ν(C≡N)] (*11*), which is a molecular tag that we had intentionally installed on the amine

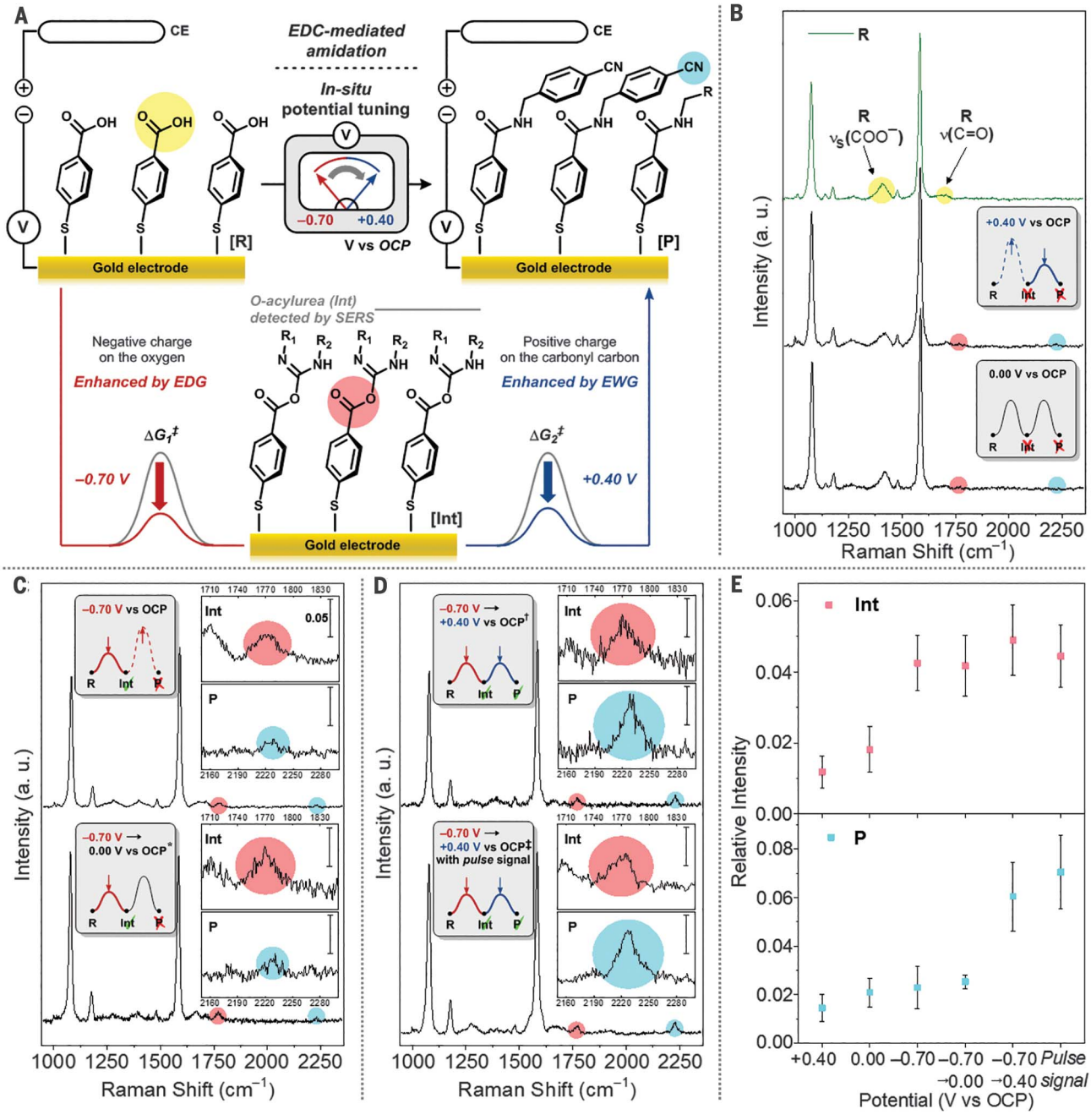


Fig. 4. Carbodiimide-mediated amidation reaction. (**A**) Schematic representation for EDC-mediated amidation of 4-mercaptobenzoic acid. After immobilizing the 4-mercaptobenzoic acid on the gold electrode, EDC (0.3 M), 4-(aminomethyl)benzonitrile (2 mM), and $NaClO_4$ (1 M) were added to MES-buffered (0.1 M) aqueous solution (pH 5.5), and reactions were run at 25°C for 1 hour. $\Delta G^{\ddagger}$, Gibbs free-energy barrier. (**B**) Representative spectra of 4-mercaptobenzoic acid (top graph, green line, R), EDC-mediated amidation with +0.40 V versus OCP (middle graph), and OCP (bottom graph). (**C**) Representative SERS spectra of EDC-mediated amidation with −0.70 V (top graph) and a sequence of −0.70 V to OCP (bottom graph). Int, intermediate. (**D**) Representative SERS spectra of EDC-mediated amidation after a sequence of −0.70 V to +0.40 V (top graph) and pulsed potential condition (bottom graph). (**E**) Relative intensities of Int and P under different potentials. The prominent peak assigned to the vibration of the aromatic ring at ~1070 cm^{-1} was used as a reference peak. *Run at −0.70 V for 40 min, then at OCP for 20 min. †Run at −0.70 V for 40 min, then at +0.40 V for 20 min. ‡Ten cycles were conducted, where each cycle consisted of −0.70 V for 3 min, then +0.40 V for 3 min (total reaction time of 1 hour). The representative peaks of R, Int, and P are highlighted in yellow, pink, and cyan, respectively. Insets show an enlarged SERS spectrum of the highlighted regions. Scale bars in (C) and (D) indicate 0.05 relative intensity. The points and error bars in the relative intensity graph (E) indicate the means and standard deviations of four to seven experiments, respectively. MES, 2-(*N*-morpholino)ethanesulfonic acid; R_1, $-C_2H_5$; R_2, $-C_3H_6(NH(CH_3)_2)^+$.

coupling partner 4-(aminomethyl)benzonitrile for monitoring purposes. The emergence of this peak is the first indication that the two-step reaction we envisioned has taken place. Next, we applied −0.70 V for 40 min and then switched to +0.40 V for 20 min, which substantially increased the intensity of the peak at 2228 cm^{-1} (Fig. 4D), suggesting that the positive potential accelerated the second step of the reaction, as anticipated. Finally, a pulse mode was tested, where −0.70 V was applied for 3 min followed by 3 min of +0.40 V and repeated 10 times during a total reaction time of 1 hour. Both the carbonyl and cyanide peaks are clearly visible in both operational modes, as illustrated in Fig. 4D. The relative intensities of the two signals observed at different potentials are summarized in Fig. 4E and show that the application of a negative potential leads to the formation of the O-acylurea. If the negative voltage is switched off, these intermediates can react slowly with the amine in solution to afford the final product, albeit in very small amounts. The second step can be accelerated by applying a positive voltage, which lowers the barrier by increasing the electrophilicity of the immobilized intermediate.

Our studies demonstrate that electrodes can be used as functional groups with adjustable inductive effects to change the chemical reactivity of molecules that are attached to them, closely mimicking the inductive effect of conventional functional groups. This constitutes a potentially powerful new way of controlling chemical reactions, offering an alternative to preparing derivatives to install electron-donating or electron-withdrawing functional groups. Classical Hammett studies that aim to gain mechanistic insight about the electronic demand of the rate-determining step can easily be imagined to be replicated through this new technique. Another promising application is in catalysis, where the electronics of the catalyst or substrate can be changed easily. Whereas the gold electrode was convenient for the proof-of-principle studies presented above, the relatively small surface area limits the effective concentration of the immobilized material to a picomolar range, which is not useful for a large-scale application. Carbon-based electrodes offer a promising solution, and future efforts will concentrate on using such electrodes to increase the scale of the reactions that can be accelerated or inhibited by the electro-inductive effect.

REFERENCES AND NOTES

1. L. P. Hammett, *J. Am. Chem. Soc.* **59**, 96–103 (1937).
2. S. Sarkar, J. G. Patrow, M. J. Voegtle, A. K. Pennathur, J. M. Dawlaty, *J. Phys. Chem. C* **123**, 4926–4937 (2019).
3. S. G. Boxer, *J. Phys. Chem. B* **113**, 2972–2983 (2009).
4. S. D. Fried, S. Bagchi, S. G. Boxer, *Science* **346**, 1510–1514 (2014).
5. C. Vericat, M. E. Vela, G. Benitez, P. Carro, R. C. Salvarezza, *Chem. Soc. Rev.* **39**, 1805–1834 (2010).
6. B. Vaidya, J. Chen, M. D. Porter, R. J. Angelici, *Langmuir* **17**, 6569–6576 (2001).
7. J. Lahann *et al*., *Science* **299**, 371–374 (2003).
8. A. Campion, P. Kambhampati, *Chem. Soc. Rev.* **27**, 241–250 (1998).
9. The peak at ~1070 cm^{-1} is assigned to an in-plane ring breathing mode coupled with C–S stretching. We designate this peak as $\nu(C–C)_{ring}$ for convenience.
10. A. Michota, J. Bukowska, *J. Raman Spectrosc.* **34**, 21–25 (2003).
11. G. Socrates, *Infrared and Raman Characteristic Group Frequencies: Tables and Charts* (Wiley, 2004).
12. H. Wackerbarth, L. Gundrum, C. Salb, K. Christou, W. Viöl, *Appl. Opt.* **49**, 4367–4371 (2010).
13. M. J. Pelletier, *Appl. Spectrosc.* **57**, 20A–42A (2003).
14. S. E. J. Bell *et al*., *Angew. Chem. Int. Ed.* **59**, 5454–5462 (2020).
15. A. J. Bard, L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications* (Wiley, ed. 2, 2000).
16. N. Miyaura, K. Yamada, A. Suzuki, *Tetrahedron Lett.* **20**, 3437–3440 (1979).
17. Y. Zhao, L. Du, H. Li, W. Xie, J. Chen, *J. Phys. Chem. Lett.* **10**, 1286–1291 (2019).
18. The peaks at 1068 and 1079 cm^{-1} are assigned to in-plane ring breathing modes coupled with C–Br stretching and C–S stretching, respectively. We designate these peaks as $\nu(C–Br)$ and $\nu(C–C)_{ring}$ for convenience.
19. G. Varsanyi, *Vibrational Spectra of Benzene Derivatives* (Elsevier, ed. 1, 1969).
20. F. Benz *et al*., *Nano Lett.* **15**, 669–674 (2015).
21. C. Amatore, A. Jutand, G. Le Duc, *Chem. Eur. J.* **17**, 2492–2503 (2011).
22. C. Amatore, A. Jutand, G. Le Duc, *Angew. Chem. Int. Ed.* **51**, 1379–1382 (2012).
23. S. B. Kedia, M. B. Mitchell, *Org. Process Res. Dev.* **13**, 420–428 (2009).
24. J. F. Hartwig, *Inorg. Chem.* **46**, 1936–1947 (2007).
25. E. Valeur, M. Bradley, *Chem. Soc. Rev.* **38**, 606–631 (2009).
26. J. P. Wolfe, S. L. Buchwald, *Angew. Chem. Int. Ed.* **38**, 2413–2416 (1999).
27. J. Heo *et al*., Data for the paper: "Electro-inductive effect: electrodes as functional groups with tunable electronic properties". Zenodo (2020); http://doi.org/10.5281/zenodo.3970448.

ACKNOWLEDGMENTS

Funding: This research was supported by the Institute for Basic Science (IBS-R010-A1) in Korea. This work was also supported by the Basic Science Research Program (2015R1A3A2033469) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT. **Author contributions:** J.H., H.A., J.W., S.W.H., and M.-H.B. conceived and designed the project and wrote the manuscript; J.H., H.A., and J.W. carried out the experiments; J.H. and J.W. performed computational studies; J.G.S., H.K.S., and T.G.L. conducted TOF-SIMS analyses; M.-H.B. organized the research; and all authors analyzed the data, discussed the results, and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The supplementary materials contain microscope images, cyclic voltammetry studies, density functional theory calculations, SERS spectra, TOF-SIMS studies, and nuclear magnetic resonance spectra. The raw data of the base-catalyzed hydrolysis reaction before the background correction can be found at Zenodo (*27*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/214/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 and S2
References (*28*–*45*)

8 March 2020; accepted 21 August 2020
10.1126/science.abb6375

NOVEL COMMUNITIES

Increased extinction in the emergence of novel ecological communities

John M. Pandolfi[1]*†, Timothy L. Staples[1]†, Wolfgang Kiessling[2]

Environmental change is transforming ecological assemblages into new configurations, resulting in novel communities. We developed a robust methodology to detect novel communities, examine patterns of emergence, and quantify probabilities of local demographic turnover in transitions to and from novel communities. Using a global dataset of Cenozoic marine plankton communities, we found that the probability of local extinction, origination, and emigration during transitions to a novel community increased two to four times that of background community changes. Although rare, novel communities were five times more likely than chance to shift into another novel state. For marine plankton communities at a 100,000-year time grain, novel communities were sensitive to further extinctions and substantial community change.

Profound changes in the biodiversity of global ecosystems (*1*, *2*) are leading to the formation of novel communities, where species composition and diversity are transformed into new, nonhistorical configurations with altered ecosystem functions (*3*–*6*). Key factors driving novel community emergence include the rapid pace of global climate change (*7*–*9*), breakdown of biogeographic barriers, species invasions, and ecosystem degradation (*10*–*12*). Here, we introduce a reproducible, objective, and quantitative approach to detect novel communities from time series compositional data, apply this approach to investigate the frequency of emergence and transition probabilities of community novelty over long temporal scales, and explore the demographic processes that influence the rise and fall of novel community states.

The conceptual basis of novel ecosystems spans multiple disciplines. For conservation biologists, novel ecosystems contain historically unprecedented combinations of species that are driven by human agency, often with altered ecological functions (*3*). The no-analog community concept from plant paleoecology focuses on the taxonomic composition or environmental framework of past communities that are compositionally unlike any found today (*13*). A variety of quantitative analytical approaches have been used to understand how and why past (*13*–*16*) or even future (*4*) vegetation dynamics in focal ecosystems differ from the present. Early emphasis on the role of novel climates in ecological change (*13*, *17*) has prompted more recent robust quantitative investigation of the geographic variation and temporal distribution of novel climates (*8*, *16*), including under various Representative Concentration Pathway scenarios (*18*, *19*).

We build on these previous approaches to provide a new novelty-detection framework (*20*) that accords with modern definitions of novel ecosystems (*3*) and enables comparative ecosystem approaches to understand general trends, causes, and consequences of novelty on a global scale. Our framework defines novel communities as having two signals of compositional change: (i) rapid change to (ii) an unprecedented state. We quantified these by decomposing an ecological time series within a site into two sets of dissimilarity scores (using the Jaccard index): a metric of community change through time (rate of change, called "instantaneous dissimilarity") (Fig. 1, A and D) and the smallest dissimilarity to any past composition (community novelness, called "cumulative dissimilarity") (Fig. 1, A and C). We classified communities as novel only where these observed dissimilarities exceeded expectations, generated for each time series using parametric spline-based models (*20*). These models estimated average dissimilarities (Fig. 1B, dashed lines) that could increase or decrease along the time series, meaning that novelty was assessed relative to local windows of time. Novelty occurred where observed dissimilarities exceeded 95% predictive intervals of expectations (Fig. 1B, gray shading). This resulted in four possible community states: faster-than-expected community turnover ("instantaneous novelty," I), a community state more distinct than expected from any prior state ("cumulative novelty," C), coincident instantaneous and cumulative novelty ("novel community," N), and "background communities" (B) that lacked novelty (Fig. 1B) (*20*).

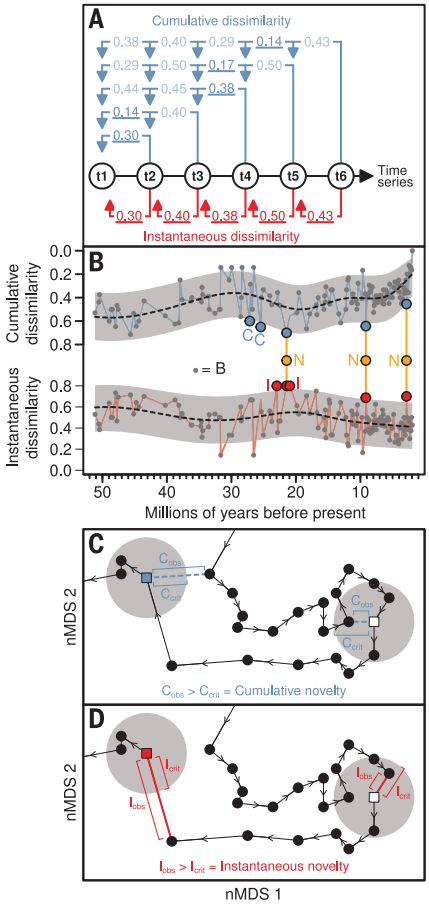


Fig. 1. Description of novel community detection, showing the time-ordered, space-restricted nature of our framework. (**A**) Calculation of instantaneous (pairwise dissimilarities; red) and cumulative (smallest dissimilarity to any past state; blue) dissimilarity in a compositional time series. (**B**) Mean expected dissimilarities (dashed lines) obtained from generalized additive models for a single time series. Observed dissimilarities (red and blue lines) were calculated for 100,000-year sampling bins. Gray shading indicates the upper and lower 95% predictive boundary. Bins that exceeded the upper boundary were classified as either instantaneous novelty (I, in red) or cumulative novelty (C, in blue; *y* axis is inverted); bins classified as both were considered to be true novel communities (N, in orange); and bins not classified as I, C, or N are background states (B, in gray). (**C** and **D**) Examples of (C) cumulative and (D) instantaneous novelty test projected on ordination axes. Points are time series sampling bins (in direction of arrows); square points indicate two example target compositions, but novelty framework tests were applied to all time series points. Target points are colored where observed dissimilarities (C_{obs} and I_{obs}, respectively) exceed 95% predictive intervals (radii of gray circles and C_{crit} and I_{crit}, respectively). Gray circle radii are equivalent to the upper edge of grey predictive regions in (B).

[1]Australian Research Council Centre of Excellence for Coral Reef Studies, School of Biological Sciences, The University of Queensland, St Lucia, Queensland 4072, Australia.
[2]GeoZentrum Nordbayern, Friedrich-Alexander Universität (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany.
***Corresponding author. Email: j.pandolfi@uq.edu.au**
†These authors contributed equally to this work.

We used our novelty-detection framework to investigate the emergence probability and demographic drivers of community novelty in the Cenozoic marine plankton record, using the Neptune Sandbox (NSB), a global set of microfossil data from deep sea drilling cores (*21*, *22*). Incorporating updated taxonomy and age models allowed us to build community data for groups of taxa across geological time. Each time series in our analyses consisted of single or multiple cores aggregated into Longhurst biogeographical provinces (*23*), with species presence and absence grouped separately for four groups of marine plankton (calcareous nannoplankton, foraminifers, radiolarians, and diatoms) every 100,000 years over the past 66 million years (fig. S1).

We found emergence probabilities of novelty to be consistent across the four marine planktonic groups. These probabilities estimated the frequency that novel planktonic communities emerged over 100,000-year time grains and under natural marine conditions, in the absence of human impacts (Fig. 2). These probabilities were sensitive to the α cut-off threshold we used to detect novelty, which was set to 0.05 (fig. S2A), but not choice of dissimilarity index (fig. S3). However, the proportional overlap among the three novelty classifications was largely stable (fig. S2, B and C), and there was a strong positive Pearson correlation between instantaneous and cumulative dissimilarity scores [correlation coefficient (r) = 0.637; 95% confidence interval (CIs), 0.628 to 0.646; n = 17,719 sampling bin comparisons; $P \leq 0.001$]. Thus, subsequent analyses built on particular sets of novel communities are unlikely to be an artifact of our arbitrary α cut-off.

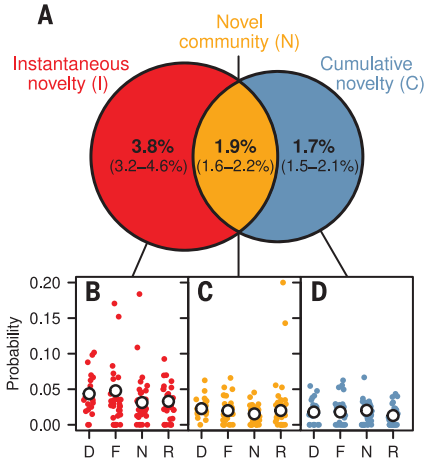


Fig. 2. Novelty emergence probabilities for four marine planktonic groups. (**A**) Venn diagram with overall mean and 95% CI probabilities of novelty emergence (table S1). (**B** to **D**) Novelty emergence probabilities and raw proportions for individual planktonic groups (tables S2 to S4). Colored points indicate raw proportion of sampling bins in each time series categorized as each novelty classification. White outlined points are mean probabilities for each planktonic group (D, diatoms; F, foraminifers; N, calcareous nannoplankton; and R, radiolarians). Lines indicating 95% CIs are obscured behind mean estimates.

We also compared the observed pair-wise probabilities of community state transitions with nonautocorrelated estimates calculated from the occurrence probabilities of each state in the transition multiplied together (*20*). Most observed transitions were from background to background states (Fig. 3, right). Transitions between the three types of novelty and a background state were much rarer (Fig. 3, middle), and many occurred less often than expected by chance. Transitions from one novelty state to another were rarer still and, with the exception of one transition, led to subsequent novelty between 2 and 10 times more often than expected (Fig. 3, left). Thus, despite the infrequent occurrence of novelty throughout our time series, transitions between novelty states were disproportionately followed by subsequent novelty states, demonstrating a heightened propensity for novelty to beget further novelty.

We evaluated the role of four demographic processes in driving transitions between one community state and another (*20*). Local extinction (the permanent loss of a species from a time series) and emigration (a transient loss with the species reappearing later in the time series) lead to species loss, whereas local origination (the first occurrence of a species in a time series) and immigration (a species subsequently reappearing after a temporary disappearance) lead to species gain. Transitions from background to background states acted as a reference baseline for comparing demographic probabilities.

Species turnover was a major feature in the transition to novelty (fig. S4). On average, taxa in the preceding community had a 10 to 14% probability of going locally extinct in the transition to a novel community, which is more than twice as high as transitions to background communities (Fig. 4A). Similarly, taxa present in the newly emerged novel community had an 18 to 27% probability on average of being entirely new to the time series (Fig. 4B, local origination), which was more than three times higher than transitions to background states. Species gain in the transition to instantaneous novelty (I) was driven more by species recolonization (Fig. 4B, immigration) than by the origination of new taxa, which is consistent with instantaneous novelty being a large shift to a state similar to past states. By contrast, species gain in transitions to new, previously unseen community states (C) and novel communities (N) came about through the addition of new taxa, likely through invasion or speciation.

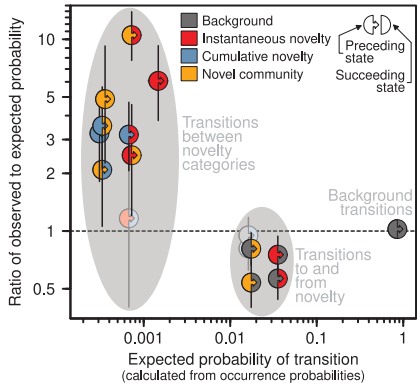


Fig. 3. Expected probability of transitions between each novelty classification, and the ratio of observed to expected transition probabilities. Points are halved and dual-colored, representing the classification of the preceding and succeeding community. *y* axis points are ratios. Faded points are transitions where the observed-to-expected ratio was not different from one. Both axes are on a natural log scale. The result summary is available in table S5.

In contrast to the emergence of novelty, taxonomic turnover played a relatively minor role in the subsequent transition from novelty (I, C, or N) to background states (Fig. 4, A and B). However, when novel communities transition to a subsequent, different novel state, these transitions were accompanied by increased probabilities of demographic changes, including local extinction (Fig. 4A) and local origination (Fig. 4B). Pairing this result with the increased probability of novel communities emerging after existing novel communities (Fig. 3, left), naturally occurring novel communities recorded from the deep-sea marine record had the potential to cascade through multiple novelty transitions, each associated with increased rates of extinction and origination.

To evaluate the robustness of our results, we computed sample completeness (figs. S5D and S6); plotted species accumulation curves (fig. S7); undertook simulations (figs. S8 and S9 and table S12); and accounted for position within time series (figs. S10 and S11), potential reworking (figs. S12 to S14), and species richness (figs. S5C and S15). We also conducted additional comparative analyses, with novelty assessed at the scale of individual cores (figs. S16 to S18), with low richness communities excluded (figs. S19 to S21), with varying sampling bin widths (200,000 to 500,000 years) (figs. S22 to S24), and analyzing taxa separately (figs. S25 and S26) and through time (fig. S5, A and B) (*20*). Our results were largely consistent across these comparative scenarios.

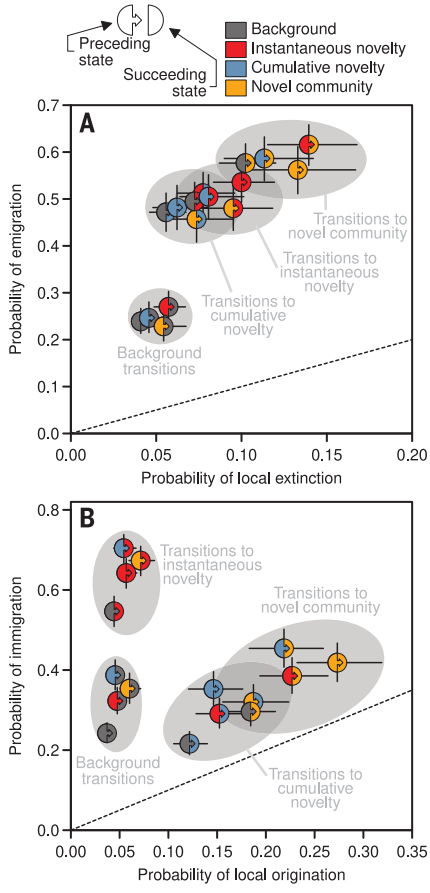


Fig. 4. Mean probabilities of taxonomic loss and gain in the transition between two communities along a time series. (**A**) Probability of local extinction plotted against probability of emigration. (**B**) Probability of local origination plotted against probability of immigration. Points are halved and dual-colored, representing the classification of the preceding and succeeding community. The dashed line represents a 1:1 ratio; error bars are 95% CIs. Model summaries are available in tables S6 to S11.

Our approach provides a quantitative tool with which to examine the frequency and drivers of novelty in ecological communities that can be broadly applied at any temporal scale, for any taxonomic group, and any ecosystem. However, we caution that additional factors such as generation times and the time frame under study must be considered in comparative analyses of the frequency or drivers of community novelty. For example, long-term studies that connect ecological and evolutionary time scales need to consider the role of speciation (and extinction) in ecosystems from the deep past when comparing the frequency of novelty with present ecosystems. Our analyses of novelty in marine plankton communities across 66 million years inevitably included an evolutionary component that could lead to higher estimates of novelty in fossil communities relative to modern communities. However, short-term fluctuations in community structure would have been averaged out over the 100,000-year time bins analyzed in our fossil time series data, potentially leading to lower estimates of past novelty than in modern communities sampled over much shorter time intervals (*24*). These and other factors will need to be assessed before fossil assemblages can be used as an ecological baseline for the frequency of novelty in living ecosystems. Nevertheless, our analyses provide a foundation for understanding the past frequency and demographic drivers of novelty in natural settings over an extended time frame and in the absence of human impacts.

Of particular concern to conservation biologists is the fate of modern communities that have transitioned into novel states (*25*): Will these new configurations of species persist, or are they likely to lead to further community change or even ecosystem collapse? Novelty in our fossil marine planktonic communities most often transitioned to a background state, suggesting that many shifts to novel states may persist through time. Where functional redundancy exists, these changes in species composition may not reduce ecosystem functioning per se. However, deep time transitions to novel communities in marine plankton exhibited both species loss and species gain but ultimately resulted in compositional shifts with at least twice the probability of local extinction (Fig. 4A). We also observed that novel communities were five times more likely to give rise to other novel communities than expected (Fig. 3, left). Although there are fundamental differences in ecological and environmental factors governing species composition over different time scales, our results from fossil communities raise the intriguing hypothesis that modern novel ecological communities could also be more likely than expected to experience transitions to subsequent novel community states, driven in part by further extinctions. Our novelty framework is well suited to further study of modern time series to test this directly.

Although we are aware of the dangers in direct comparison of ecological dynamics over vastly different time scales, our results raise the possibility that efforts to reduce extinction risk are consistent with the active management and conservation of novelty in modern ecological communities (*26*–*28*). Under the influence of human impacts, failing to improve the conditions that brought about the transition to novelty may facilitate further novelty accompanied by additional species extinction. Our results contravene any notion that a fixed historical baseline of a community can or should be the only conservation goal for novel communities (*3*). Rather, ecosystem management should also focus on how to prevent transitions to additional previously unseen ecosystem states that are linked with heightened extinction risk.

REFERENCES AND NOTES

1. M. Dornelas *et al.*, *Science* **344**, 296–299 (2014).
2. S. A. Blowes *et al.*, *Science* **366**, 339–345 (2019).
3. R. J. Hobbs, E. Higgs, J. A. Harris, *Trends Ecol. Evol.* **24**, 599–605 (2009).
4. V. C. Radeloff *et al.*, *Ecol. Appl.* **25**, 2051–2068 (2015).
5. A. E. Magurran, *Science* **351**, 448–449 (2016).
6. R. J. Hobbs, L. E. Valentine, R. J. Standish, S. T. Jackson, *Trends Ecol. Evol.* **33**, 116–128 (2018).
7. D. Stralberg *et al.*, *PLOS ONE* **4**, e6825–e6825 (2009).
8. A. Ordonez, J. W. Williams, J. C. Svenning, *Nat. Clim. Chang.* **6**, 1104–1109 (2016).
9. B. R. Scheffers *et al.*, *Science* **354**, aaf7671 (2016).
10. A. E. Lugo, T. A. Carlo, J. M. Wunderle Jr., *Anim. Conserv.* **15**, 233–241 (2012).
11. A. Traveset *et al.*, *Proc. Biol. Sci.* **280**, 20123040 (2013).
12. A. Anton *et al.*, *Nat. Ecol. Evol.* **3**, 787–800 (2019).
13. J. W. Williams, B. N. Shuman, T. Webb III, *Ecology* **82**, 3346–3362 (2001).
14. W. Finsinger, T. Giesecke, S. Brewer, M. Leydet, *Ecol. Lett.* **20**, 336–346 (2017).
15. V. Iglesias, C. Whitlock, T. R. Krause, R. G. Baker, *J. Biogeogr.* **45**, 1768–1780 (2018).
16. K. D. Burke *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190218 (2019).
17. J. W. Williams, S. T. Jackson, J. E. Kutzbach, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5738–5742 (2007).
18. M. C. Fitzpatrick *et al.*, *Glob. Change Biol.* **24**, 3575–3586 (2018).
19. C. R. Mahony, A. J. Cannon, T. Wang, S. N. Aitken, *Glob. Change Biol.* **23**, 3934–3955 (2017).
20. Materials and methods and comparative analyses are available as supplementary materials.
21. D. Lazarus, *Math. Geol.* **26**, 817–832 (1994).
22. C. Spencer-Cervato, *Palaeontol. Electronica* **1999**, 2 (1999).
23. A. R. Longhurst, *Ecological Geography of the Sea* (Academic Press, ed. 2., 2007).
24. D. B. Kemp, K. Eichenseer, W. Kiessling, *Nat. Commun.* **6**, 8890 (2015).
25. E. Marris, *Nature* **460**, 450–453 (2009).
26. A. M. Truitt *et al.*, *Environ. Manage.* **55**, 1217–1226 (2015).
27. S. J. Capon, G. J. Palmer, *Solutions* **9**, 3 (2018).
28. S. Clement, R. J. Standish, *J. Environ. Manage.* **208**, 36–45 (2018).
29. T. Staples, TimothyStaples/novelty-cenozoic-microplankton: Primary release of code used for data analysis and modeling. Zenodo (2020); doi:10.5281/zenodo.4031861.

ACKNOWLEDGMENTS

We are grateful for discussion about this work with S. Jackson, D. Lazarus, J. Levine, and R. Norris. **Funding:** This work was funded by the Australian Research Council's Centre of Excellence for Coral Reef Studies (CE140100020) and by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation, KI 806/16-1). **Author contributions:** The ideas for the paper were conceived by J.M.P. and W.K. and enhanced by T.L.S.; T.L.S. undertook all analyses and developed analytical approaches, with input from J.M.P. and W.K.; and J.M.P. and T.L.S. wrote the paper, with W.K. contributing to writing and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data used in this study are available from the Neptune Sandbox microfossil database (http://nsb-mfn-berlin.de/search). R code to conduct all analyses and produce all tables and figures are available at Zenodo (*29*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/220/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S12
References (*30*–*43*)
MDAR Reproducibility Checklist

21 February 2020; accepted 20 August 2020
10.1126/science.abb3996

ELECTRON MICROSCOPY

Cryo-EM with sub–1 Å specimen movement

Katerina Naydenova[1], Peipei Jia[1,2]*, Christopher J. Russo[1]†

Most information loss in cryogenic electron microscopy (cryo-EM) stems from particle movement during imaging, which remains poorly understood. We show that this movement is caused by buckling and subsequent deformation of the suspended ice, with a threshold that depends directly on the shape of the frozen water layer set by the support foil. We describe a specimen support design that eliminates buckling and reduces electron beam–induced particle movement to less than 1 angstrom. The design allows precise foil tracking during imaging with high-speed detectors, thereby lessening demands on cryostage precision and stability. It includes a maximal density of holes, which increases throughput in automated cryo-EM without degrading data quality. Movement-free imaging allows extrapolation to a three-dimensional map of the specimen at zero electron exposure, before the onset of radiation damage.

Despite the current success of cryo-EM in determining the structure of numerous macromolecular complexes that were previously intractable, several outstanding problems remain: Specimen movement at the beginning of electron beam irradiation degrades image quality and reduces information about the undamaged structure (*1*). Thus, all current structures determined by cryo-EM have varying degrees of radiation damage incorporated into the resulting atomic models.

The throughput of modern synchrotron crystallography beamlines vastly exceeds that of current state-of-the-art electron microscopes; hence, high-resolution structure determination by cryo-EM is feasible but throughput is comparatively limited in practice (*2*). Rapid structure determination by cryo-EM is further hampered by structural heterogeneity in many samples, as well as by irreproducible and destructive interactions with surfaces during specimen preparation (*3*). Here, we present a physical theory of the causes of movement of cryo-EM specimens during imaging and use this work to design and implement a specimen support that eliminates movement and helps to address some of these limitations.

Proteins rapidly frozen for cryo-EM in amorphous ice are inherently low-contrast, making them difficult to track and image. Previous work using virus particle tracking on amorphous carbon supports showed a "drum-like motion" that included deformation of the support foil and the ice layer, which together cause a rotation and vertical displacement of the particles by 150 to 250 Å (*4*, *5*). More recent tracking methods yield similar types of trajectories for various smaller protein specimens (*6*, *7*). Gold nanoparticles embedded in vitreous ice provide a high-contrast fiducial marker to measure the movement of the ice during a typical high-resolution, low-dose micrograph (Fig. 1A and fig. S1). The types of movement that occur during cryo-EM imaging, ordered by magnitude, are (i) stage drift, (ii) bending of the support foil, (iii) ice bending, (iv) Brownian movement during irradiation, and (v) molecular vibrations (supplementary text). To isolate the movement of the ice layer, we eliminate the movement of the specimen support by using ultrastable gold foils that do not move during irradiation (*8*) and include the edge of the gold hole in the image for accurate (<0.1 Å) stage drift correction. What remains is then the movement of the ice within the hole and the individual particles within the ice.

From measurements of the movement of ~15,000 particles from hundreds of holes in ultrastable gold foils (*8*), with hole diameters from the typical 1 to 2 μm in standard specimen support foils down to 200 nm, a pattern of movement becomes clear (Fig. 1, B and C, and figs. S2 to S8). There is an abrupt, spatially correlated displacement of the particles at the onset of irradiation (i.e., during the first ~4 $e^-/Å^2$). This is followed by a less correlated movement that continues with reduced speed to the end of the exposure. The initial movement is particularly apparent in the tilted micrographs of the 2-μm holes, where the vertical (out of plane) displacement of the ice is 10 to 50 Å in a 30 $e^-/Å^2$ exposure (fig. S4). In contrast, the movement of the particles in 300- and 200-nm holes is isotropic (the same in the plane of the support and perpendicular to it), is spatially uncorrelated, and scales with the square root of the incident electron fluence, as would be the case for diffusional motion (figs. S2 and S9 to S12).

To explain this movement, we suggest a physical model, based on considering the stresses that are created in the specimen during vitrification and the pseudo-diffusion of the water molecules under electron irradiation (*9*) (Fig. 2 and supplementary text). In typical cryo-EM specimens, the ice deforms twice, once during freezing and again during irradiation with the electron beam. Vitreous water (low-density amorphous ice) has 6% lower density than liquid water at 4°C (*10*, *11*). During cryo-plunging, freezing occurs over a time interval of $\leq 10^{-4}$ s (*10*). Within this time, the water density change is most rapid near the homogeneous nucleation temperature, ~235 K (*12*, *13*). As the water solidifies, the rapid cooling does not allow sufficient time for structural rearrangements of the water molecules and causes the buildup of compressive strain within the thin film (Fig. 2A). Once the compression exceeds a critical value, the film buckles, thus momentarily relieving the radial stress in the layer. Buckling only occurs if the stress exceeds a critical point, which is determined by the dimensions of the ice film, its elastic moduli, specific volume change relative to the support, and the constraints at the edge of the hole (supplementary text and figs. S13 to S15). As the vitreous film continues to cool to the temperature of the surrounding cryogen (typically liquid ethane at 90 to 93 K), more stress builds up as a result of further changes in relative density. This stress is stored in the film indefinitely at 77 K (liquid nitrogen temperature).

High-energy electron irradiation increases the effective diffusivity of water molecules by 46 orders of magnitude (Fig. 2B) (*14*), causing the ice in the electron beam to behave as an ultraviscous fluid. This change in the mechanical properties of the film results in unbalanced bending moments in it; these cause a correlated and rapid displacement of the particles in the same direction as the preexisting buckling (Fig. 1D), apparent in cryomicrographs. The movement is enabled and accompanied by random movement due to the continuous pseudo-diffusion of the surrounding water and only occurs during electron irradiation. The predicted critical aspect ratio (diameter:thickness), above which the film is expected to buckle during freezing and become unstable during irradiation, is ~11:1 (supplementary text). When the aspect ratio of the vitreous ice film is below the critical value (i.e., holes 330 nm or 180 nm in diameter for 300-Å-thick ice in Fig. 1, B and C), the ice never builds up enough stress to buckle during freezing or to deform under irradiation. Such films remain stable throughout irradiation, and only diffusive movement occurs (fig. S16); this is limited to <1 Å RMS (root mean squared) in 30 $e^-/Å^2$.

On the basis of the above theory, we propose a simple method for choosing the dimensions of a support foil to prevent buckling for a given

[1]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [2]ARC Centre of Excellence for Nanoscale BioPhotonics (CNBP), Institute for Photonics and Advanced Sensing, School of Physical Sciences, University of Adelaide, Adelaide 5005, Australia.
*Present address: Shenzhen Topmembranes Technology Co. Ltd., Shenzhen 518000, China.
†Corresponding author. Email: crusso@mrc-lmb.cam.ac.uk

specimen: The optimal foil thickness is set by the desired ice thickness, and the optimal foil hole diameter is ≤11 times the ice thickness (fig. S17). For example, for a 200-Å-thick particle in 300-Å-thick ice, the optimal hole diameter is below 330 nm. For 100-Å-thick ice, potentially desirable for sub–100-kDa specimens, the optimal diameter is below 110 nm. In such small holes, the thickness of the suspended ice is largely controlled by the foil thickness; uniform and stable thin films of vitreous water can be formed. Unfortunately, for most single-particle specimens, the desired foil hole size is less than the practical limit of conventional photolithography techniques currently used to make specimen supports, and evaporated foils of gold less than ~400 Å thick are not stable because of their polycrystalline grain structure (*15*). We propose a specimen support design, "HexAuFoil" (Fig. 3)—which entirely eliminates the buckling of the suspended amorphous ice during irradiation—and provide a scaleable method for manufacturing these supports based on phase interference lithography [Talbot

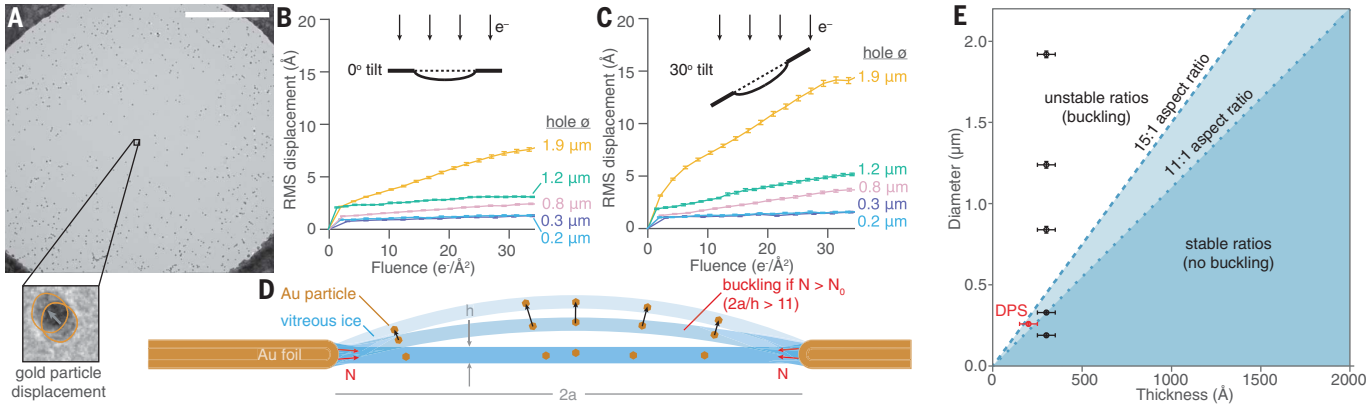
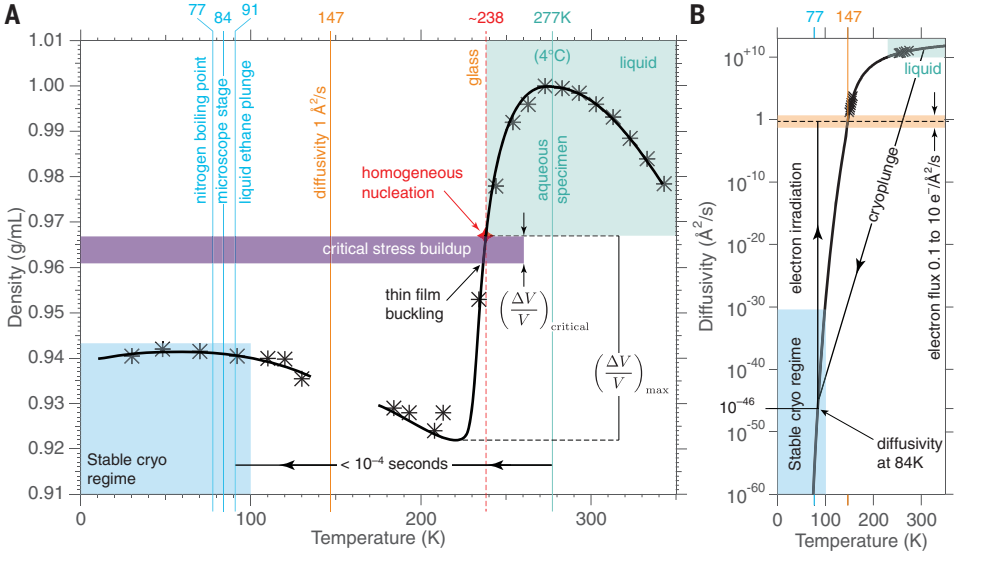


Fig. 1. Movement of gold nanoparticles in vitreous ice in a range of foil hole sizes. (**A**) Typical drift-corrected electron micrograph used for tracking gold particles in vitreous water on all-gold supports; scale bar, 0.5 µm. The inset (20 nm × 20 nm) shows an overlay of the positions of a gold nanoparticle at the beginning of irradiation and after a fluence of 60 $e^-/Å^2$. (**B** and **C**) RMS displacements of 200 to 2000 particles from 10 to 50 movies of holes of different diameters (yellow, UltrAuFoil R2/2, 1.9-µm holes; green, UltrAuFoil R1.2/1.3, 1.2-µm holes; pink, UltrAuFoil R 0.6/1, 0.8-µm holes; purple and blue, custom-made grids with 0.3-µm and 0.2-µm holes, respectively) are plotted as a function of cumulative electron fluence for 0° tilt (B) and 30° tilt (C). Error bars denote SEM. (**D**) Thin films of ice used in cryo-EM buckle during vitrification if the compressive stress (N) exceeds a critical value (N_0), determined by the aspect ratio ($2a/h$) of the film. Electron irradiation causes the film to move in response to additional stresses in it, as is evident from the correlated particle movement at the beginning of irradiation. (**E**) The range of stable hole diameters can be determined for a given ice thickness (blue shaded region). The two critical aspect ratios (dashed and dotted lines) are calculated according to two theoretical models of buckling during vitrification (supplementary text). The black and red data points show the hole sizes and ice thicknesses for the gold particle and DPS datasets, respectively. Error bars denote SD.

Fig. 2. Model of stress accumulation in thin films of amorphous ice during cryoplunging, and their response to electron irradiation. (**A**) The density of liquid and amorphous water is plotted (black markers) as a function of temperature, as reported in (*12*, *13*), with a black line to guide the eye. During cryoplunging into liquid ethane, water is rapidly cooled, typically from 277 K to 91 K (black arrow). The largest specific volume change experienced by water below its homogeneous nucleation point (red) is $(\Delta V/V)_{max}$ ≈ 6%. The thin film can only withstand compression of up to $(\Delta V/V)_{critical}$ before it buckles (purple range corresponds to a 300-Å-thick layer in a 1-µm hole). (**B**) The diffusivity of water molecules in liquid and amorphous ice is plotted (black crosses) as a function of temperature, as reported in (*14*). The black line is a fit to these values, as proposed in (*14*). The extrapolated diffusivity in amorphous ice at 84 K is vanishingly low, ~10^{-46} $Å^2/s$. The blue shaded region indicates the range of diffusivity in amorphous ice at temperatures in the 0 to 100 K range, where it is stable indefinitely. During imaging with 300-keV electrons, water molecules move pseudo-diffusively by 1 $Å^2/(e^-/Å^2)$ (*9*). At typical imaging fluxes of 0.1 to 10 $(e^-/Å^2)/s$, this is equivalent to 0.1 to 10 $Å^2/s$ (orange) and corresponds to an instantaneous local temperature of 147 K.



displacement (*16*, *17*)] and low-temperature evaporation (figs. S18 and S19). The holes are arranged in a hexagonal pattern in a foil whose thickness (280 Å for the example in Fig. 3F) can be matched to the specimen; the holes still remain round (fig. S19B). Interestingly, the nanoscale dimensions of the array imply that the foil has plasmon resonances in the visible range, which causes the foil to appear yellow on reflection with white light but blue on transmission (Fig. 3, A and B, and fig. S18G)—a property that might be used in the future for characterizing the specimen before imaging with electrons. The foil is suspended across a 3-mm hexagonal mesh grid; together these hexagonal arrays increase the usable area by a factor of 10 relative to a standard cryo-EM grid. More than 800 images can be acquired from a single stage position, enabling a faster rate of automated data collection (*18*, *19*), and more than 5000 individual holes can be imaged in a single 25-μm-wide hexagon (Fig. 3C). Having the edge of the gold foil in each image (Fig. 3, E and F, and fig. S20) has the additional benefit of allowing rapid and accurate drift tracking independent of the signal from the specimen within the holes, which reduces the demands on stage precision and stability.

To demonstrate the use of movement-free specimen supports for high-resolution cryo-EM, we determined the structure of the 223-kDa DNA protection during starvation (DPS) protein (*20*), plunge-frozen on grids with 280-Å-thick gold foil with 260-nm holes (Fig. 3F and figs. S20 to S22). The average resolution, from an initial reconstruction from ~9 hours of automated data collection on a modern 300-keV microscope, easily reached <2 Å; the total particle displacement was 0.86 Å RMS in 35 $e^-/Å^2$ of irradiation (Fig. 4A, fig. S23, and supplementary text). The absence of buckling also ensured that no tilting of the particles occurred during imaging. In contrast to all previous sub–2 Å resolution single-particle cryo-EM datasets to date (supplementary text), maps reconstructed from each frame showed that the first frame (1 $e^-/Å^2$ or 3 MGy) contained the most structural information (Fig. 4A) and that the quality (*B*-factor) of sequential frames decayed linearly with dose/fluence. A linear decay in *B*-factor with dose is expected from studies of radiation damage in x-ray and electron crystallography (*21*, *22*) but had never been observed for single-particle cryo-EM because of movement of the specimen at the onset of irradiation.

Current cryo-EM data processing and reconstruction algorithms derive the final reconstructed map by summing the information in each frame, down-weighted to account for movement and damage (*23*, *24*). In contrast, decoupling specimen movement from radiation damage affords new approaches to reconstruction based on the physical theory of how the structure factors decay with exposure due to mass loss and radiation damage only (supplementary text). First, a reconstruction can be calculated from a single frame corresponding to a total fluence of 1 $e^-/Å^2$ (supplementary text and fig. S24), which is less than the typical dose limit used in x-ray crystallography [10 MGy (*25*)]. Second, the complex structure factors at each pixel in Fourier space can be fit with exponential functions decaying with dose, which extrapolate back to an initial "undamaged" value at zero exposure (Fig. 4, B and C). A similar method for amplitudes only was previously proposed but has not been widely used in x-ray crystallography (*26*). Maps produced by both of these methods show improved densities for radiation-sensitive side chains, ordered water molecules, and other complexed ions; high-resolution, <10 MGy reconstructions avoid potential problems in accurately modeling radiation-sensitive parts of biological molecules and make the per-atom *B*-factors directly interpretable as movement. Further, the progressive effects of radiation are directly evident in the sequential maps and can be used to improve atomic modeling techniques and inform understanding of radiation damage (Fig. 4D). After fully removing movement, it is also possible to separate out the effects of other factors that reduce resolution, such as particle heterogeneity and deviation from symmetry (fig. S22A). Elimination of beam-induced foil and ice movement provides an overall *B*-factor improvement of 54 ± 10 $Å^2$ over motion-corrected micrographs for this specimen in the sub–4 Å resolution range (fig. S22B).

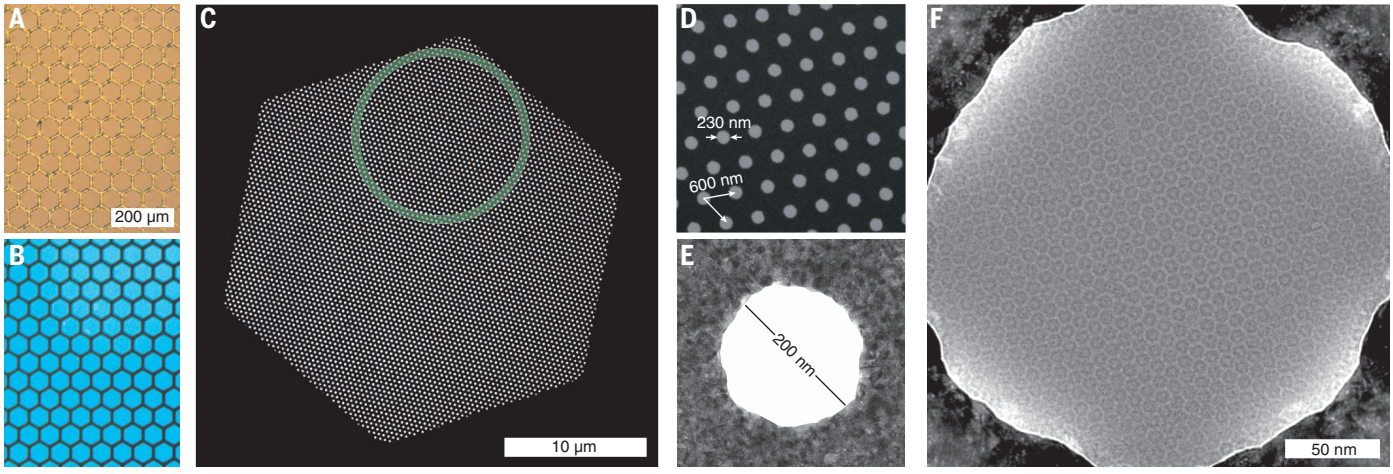


Fig. 3. HexAuFoil is an all-gold specimen support designed for movement-free cryo-EM imaging. (**A** and **B**) Optical micrographs showing reflected (A) and transmitted (B) unpolarized white illumination of the patterned gold foil (hexagonal array of 200-nm-diameter holes with 600-nm pitch) on a 600-mesh thin-bar gold grid. The scale and the corresponding area are the same for (A) and (B). The blue color of the foil in transmitted light is due to a strong red absorption enhancement by the periodic hole pattern (fig. S18G). (**C**) Transmission electron micrograph of a single hexagonal grid opening on one of these grids. A 3-mm grid contains ~800 of these hexagons, each of which includes more than 5000 holes in a regular pattern. The green circle encloses more than 800 holes, which can all be imaged at high magnification without moving the stage during high-speed data collection (fig. S21). (**D**) Transmission electron micrograph of the holey gold foil. The arrows show the pitch of the regular hexagonal pattern. (**E**) Transmission electron micrograph of a single hole in the nanocrystalline foil. The roundness of the 200-nm hole is improved by reducing the gold grain size to 10 nm or less, which can be achieved by deposition onto a cooled template (fig. S19). (**F**) Low-dose transmission electron micrograph, at 1.5-μm defocus, of the protein DPS (223 kDa) vitrified on a HexAuFoil grid with 260-nm holes.

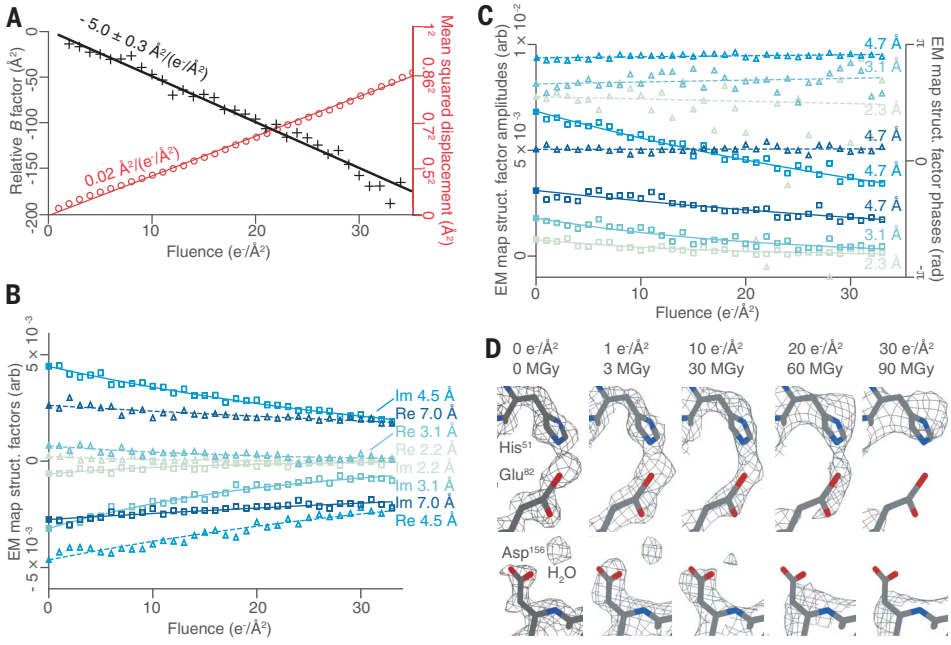


Fig. 4. Structure of DPS determined at <2 Å resolution by extrapolation to zero dose with the use of a 260-nm hole support. (**A**) Plot of the mean squared displacement during irradiation (red) for all DPS particles used in the reconstruction, and the relative *B*-factor for each frame with respect to the first (black) with linear fits to both. The displacement of the particles corresponds to diffusion with a constant of 0.02 $Å^2/(e^-/Å^2)$ (red line). The *B*-factor decay agrees with the expected slope from radiation damage alone (*32*). (**B**) The real (triangles) and imaginary (squares) parts of selected Fourier pixels are plotted as a function of fluence. (**C**) The phases (triangles) and amplitudes (squares) of selected Fourier pixels (at 2.3 Å, 3.1 Å, and two different pixels at 4.7 Å resolution) are plotted as a function of fluence. The real and imaginary parts (B) or phases and amplitudes (C) are extrapolated to their values before the onset of irradiation, corresponding to the undamaged structure (solid symbols at 0 fluence). The lines in (B) and (C) are exponential fits (to the real parts, imaginary parts, and amplitudes) or linear fits (to the phases). (**D**) Selected side chains and a water molecule from zero-dose extrapolated and per-frame DPS reconstructions show the progression of radiation damage. The residues from the refined model are colored by atom (gray, C; blue, N; red, O), and the contoured density map is shown as a mesh.

The specimen support described here reduces particle movement in a cryomicrograph to the limit set by pseudo-diffusion, which is on average less than the diameter of a hydrogen atom. This allows reconstruction of a complete map of an undamaged structure at 1.9 Å resolution. We note that interaction of the specimen with the air/water interface remains a limitation for cryo-EM specimen preparation (*3*) and can be the dominant factor determining the success of a project (fig. S22D). Several approaches have been used to address this, including addition of a surface such as functionalized graphene (*27*–*30*) and decreasing the time of interaction with surfaces before freezing (*31*). Both are compatible with the described support, and current work is focused on integrating the small aspect ratio and hexagonal pattern of holes with other existing technologies and manufacturing the devices at scale. Movement-free imaging will also allow the investigation of lower-temperature cryomicroscopy (closer to 0 K) where the secondary effects of radiation damage may be further reduced from those at liquid nitrogen temperatures, thus affording more contrast per image. With these improvements, cryo-EM will continue to rapidly expand our understanding of the structures and functions of biological molecules.

REFERENCES AND NOTES

1. R. Henderson, C. J. Russo, *Microsc. Microanal.* **25** (suppl. 2), 4–5 (2019).
2. M. Saur *et al.*, *Drug Discov. Today* **25**, 485–490 (2020).
3. R. M. Glaeser, *Curr. Opin. Colloid Interface Sci.* **34**, 1–8 (2018).
4. R. Henderson *et al.*, *J. Mol. Biol.* **413**, 1028–1046 (2011).
5. A. F. Brilot *et al.*, *J. Struct. Biol.* **177**, 630–637 (2012).
6. S. Q. Zheng *et al.*, *Nat. Methods* **14**, 331–332 (2017).
7. J. Zivanov, T. Nakane, S. H. W. Scheres, *IUCrJ* **6**, 5–17 (2019).
8. C. J. Russo, L. A. Passmore, *Science* **346**, 1377–1380 (2014).
9. G. McMullan, K. R. Vinothkumar, R. Henderson, *Ultramicroscopy* **158**, 26–32 (2015).
10. J. Dubochet *et al.*, *Q. Rev. Biophys.* **21**, 129–228 (1988).
11. T. Loerting *et al.*, *Phys. Chem. Chem. Phys.* **13**, 8783–8794 (2011).
12. F. Mallamace *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18387–18391 (2007).
13. V. Holten, M. A. Anisimov, *Sci. Rep.* **2**, 713 (2012).
14. R. S. Smith, Z. Dohnalek, G. A. Kimmel, K. P. Stevenson, B. D. Kay, *Chem. Phys.* **258**, 291–305 (2000).
15. C. J. Russo, L. A. Passmore, *J. Struct. Biol.* **193**, 33–44 (2016).
16. H. H. Solak, C. Dais, F. Clube, *Opt. Express* **19**, 10686–10691 (2011).
17. K. Jefimovs *et al.*, in *Advances in Patterning Materials and Processes XXXIV*, C. K. Hohle, Ed. (SPIE, 2017), pp. 140–146.
18. C. Suloway *et al.*, *J. Struct. Biol.* **151**, 41–60 (2005).
19. M. Schorb, I. Haberbosch, W. J. H. Hagen, Y. Schwab, D. N. Mastronarde, *Nat. Methods* **16**, 471–477 (2019).
20. R. A. Grant, D. J. Filman, S. E. Finkel, R. Kolter, J. M. Hogle, *Nat. Struct. Biol.* **5**, 294–303 (1998).
21. M. Warkentin, J. B. Hopkins, J. B. Haber, G. Blaha, R. E. Thorne, *Acta Crystallogr. D* **70**, 2890–2896 (2014).
22. J. Hattne *et al.*, *Structure* **26**, 759–766.e4 (2018).
23. S. H. W. Scheres, *eLife* **3**, e03665 (2014).
24. T. Grant, N. Grigorieff, *eLife* **4**, e06980 (2015).
25. R. Henderson, *Proc. R. Soc. B* **241**, 6–8 (1990).
26. K. Diederichs, S. McSweeney, R. B. Ravelli, *Acta Crystallogr. D* **59**, 903–909 (2003).
27. K. Naydenova, M. J. Peet, C. J. Russo, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11718–11724 (2019).
28. X. Fan *et al.*, *Nat. Commun.* **10**, 2386 (2019).
29. E. D'Imprima *et al.*, *eLife* **8**, e42747 (2019).
30. F. Wang *et al.*, *J. Struct. Biol.* **209**, 107437 (2020).
31. A. J. Noble *et al.*, *eLife* **7**, e34257 (2018).
32. M. J. Peet, R. Henderson, C. J. Russo, *Ultramicroscopy* **203**, 125–131 (2019).

ACKNOWLEDGMENTS

We thank A. Howie for discussions and advice throughout this project; M. J. Peet, R. Henderson, G. McMullan, T. Nakane, S. Scheres, R. A. Crowther, A. Leslie, J. Dickerson, L. A. Passmore, and P. A. Midgley for helpful discussions; Y. Lee and P. C. Edwards for the DPS specimen; S. Chen, G. Cannone, G. Sharov, and A. Yeates of the Laboratory of Molecular Biology Electron Microscopy Facility; K. Sader, P. Qian, and P. Castro-Hartmann (Thermo Fisher Scientific) for assistance with data collection on the Cambridge Pharma Consortium microscope housed in the Department of Materials Science and Metallurgy, University of Cambridge; and J. Grimmett and T. Darling (MRC Laboratory of Molecular Biology Scientific Computing) and the MRC Laboratory of Molecular Biology Mechanical and Electronics workshops for technical assistance. **Funding:** Supported by a Vice-Chancellor's Award (Cambridge Commonwealth, European and International Trust) and a Bradfield scholarship (K.N.), an Australian Nanotechnology Network overseas travel fellowship (P.J.), and Medical Research Council grant MC_UP_120117. **Author contributions:** K.N., P.J., and C.J.R. designed and performed the templated small-hole grid fabrication experiments; K.N. and C.J.R. collected and analyzed the cryo-EM data, performed other movement, density, and low-temperature physical vapor deposition–related experiments, developed the theory of ice movement, and wrote the paper. **Competing interests:** The authors are inventors on a patent application GB2004272.7 on the design of the specimen support, filed by the Medical Research Council as part of United Kingdom Research and Innovation. **Data and materials availability:** The entire DPS dataset is publicly available in the Electron Microscopy Public Image Archive (EMPIAR-10445). The final reconstructed maps from each frame and the weighted sum are deposited in the Electron Microscopy Data Bank (EMD-11210), and the refined atomic model in the Protein Data Bank (PDB-6ZGL). All code for zero-dose data processing is freely available from the author's website, www.mrc-lmb.cam.ac.uk/crusso/, and is deposited in Zenodo at https://dx.doi.org/10.5281/zenodo.3944689.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/223/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Table S1
References (*33*–*82*)

24 March 2020; accepted 15 July 2020
10.1126/science.abb7927

PLANT SCIENCE

WUSCHEL triggers innate antiviral immunity in plant stem cells

Haijun Wu[1], Xiaoya Qu[1], Zhicheng Dong[2], Linjie Luo[1], Chen Shao[1], Joachim Forner[3]*, Jan U. Lohmann[3], Meng Su[1], Mengchu Xu[1], Xiaobin Liu[2], Lei Zhu[4], Jian Zeng[1], Sumei Liu[1], Zhaoxia Tian[1]†, Zhong Zhao[1]†

Stem cells in plants constantly supply daughter cells to form new organs and are expected to safeguard the integrity of the cells from biological invasion. Here, we show how stem cells of the *Arabidopsis* shoot apical meristem and their nascent daughter cells suppress infection by cucumber mosaic virus (CMV). The stem cell regulator WUSCHEL responds to CMV infection and represses virus accumulation in the meristem central and peripheral zones. WUSCHEL inhibits viral protein synthesis by repressing the expression of plant *S*-adenosyl-L-methionine–dependent methyltransferases, which are involved in ribosomal RNA processing and ribosome stability. Our results reveal a conserved strategy in plants to protect stem cells against viral intrusion and provide a molecular basis for WUSCHEL-mediated broad-spectrum innate antiviral immunity in plants.

Plant stem cells and their newly differentiated daughter cells in the shoot apical meristem (SAM) provide cells for all aboveground organs during postembryonic development (*1*). This tissue is free from viral invasion. Indeed, meristem tip culture eliminates viruses from infected plants and is widely used to develop virus-free cultivars (*2*). Here, we analyzed the broad-spectrum antiviral mechanism that safeguards pluripotent cells in the SAM from viral invasion.

To test the hypothesis that the top of the SAM is virus free (*3*), we followed virus distribution in the *Arabidopsis* SAM after inoculating the leaves with cucumber mosaic virus (CMV) (Fig. 1A). At 9 days postinoculation (dpi), we observed CMV accumulations in the shoot and the rib zone of the SAM, but not in the central zone (CZ), peripheral zone (PZ), or floral primordia (Fig. 1, B to D, and figs. S1 and S2). We observed CMV accumulation just below the *WUSCHEL* (*WUS*) (*4*, *5*) expression domain (Fig. 1, E and F). In the *clv3-7* mutant that ectopically expressed *WUS* in the L2 cell layer (*6*–*8*) (Fig. 1G), CMV moved upward and remained beneath the WUS protein–containing domain (Fig. 1H and fig. S3). We then investigated whether the *WUS* expression domain blocks virus movement or if WUS inhibits virus accumulation. We generated an *mp clv3-7* double mutant with an enlarged SAM, which allowed us to inoculate CMV directly into and below the *WUS* expression domain (Fig. 1I). We performed a serial sectioning–based mismatch assay, with one section for the detection of *WUS* expression and the adjacent section for the detection of CMV. Although the virus was directly inoculated into stem cells in the L1 cell layer and the organizing center (OC) in the L2 cell layer, where the WUS protein is located, we observed that the CMV only spread under the L2 layer (Fig. 1, J and K), suggesting that the WUS protein–containing regions, including stem cells and the OC, are resistant to CMV invasion.

To test whether WUS proteins are involved in the host defense against viruses, we examined WUS protein distribution patterns during CMV infection at 9 dpi using *wus/pCLV3::mCherry-NLS*; *pWUS::WUS-GFP* rescue plants. In virus-free plants, WUS protein expression overlapped with *CLV3* transcript expression in stem cells (Fig. 2, A to D). Upon CMV invasion, WUS protein expression (Fig. 2, E, G, and H, and fig. S4), but not transcript expression (fig. S5), increased. WUS protein was ectopically expressed in the entire PZ and young floral primordia in a pattern complementary to the CMV distribution in the shoot apex (Fig. 1D).

In plants in which WUS was inducible with dexamethasone (DEX) treatment [*pUBQ10::WUS-GR* plants (*9*)], we induced WUS 1 day after inoculating with CMV. Without induction, 89% of the plants were infected by viruses.

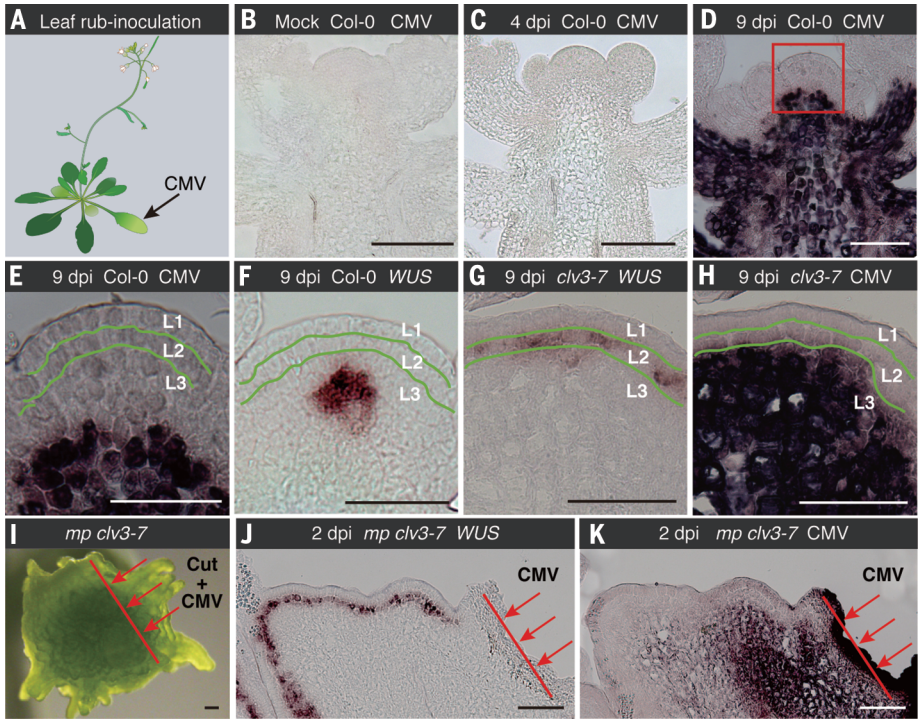


Fig. 1. Plant stem cells and their progeny are immune to CMV infection. (**A**) CMV was inoculated in leaves to trigger systemic infection in *Arabidopsis*. (**B** to **D**) CMV distributions in mock (B), 4 dpi (C), and 9 dpi (D) plants. Red box indicates the SAM. Col-0, Columbia-0. (**E** and **F**) CMV (E) and *WUS* (F) distributions in the SAM at 9 dpi. (**G** and **H**) *WUS* and CMV accumulation patterns in the *clv3-7* mutant. (**I** to **K**) *WUS* (J) and CMV (K) distributions in the *mp clv3-7* double mutant with viral infection (I). Red lines indicate the surgical incision in the SAM. Red arrows indicate the location of CMV inoculation. Scale bars, 100 μm in (B) to (D) and (I) to (K) and 50 μm in (E) to (H).

[1]Hefei National Laboratory for Physical Sciences at the Microscale, CAS Center for Excellence in Molecular Plant Sciences, School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230027, China. [2]School of Life Sciences, Guangzhou University, Guangzhou Higher Education Mega Center, Guangzhou 510006, China. [3]Department of Stem Cell Biology, Centre for Organismal Studies, Heidelberg University, Heidelberg D-69120, Germany. [4]State Key Laboratory of Biotherapy and Cancer Center, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu 610041, China.
*Present address: Max Planck Institute of Molecular Plant Physiology, Organelle Biology, Potsdam D-14476, Germany.
†Corresponding author. Email: zhzhao@ustc.edu.cn (Z.Z.); zxtian@ustc.edu.cn (Z.T.)

With DEX-induced WUS activity, 90% of the plants were free from CMV invasion (Fig. 2, I to L). To determine whether the WUS protein is required for the antiviral capacity of stem cells, we used *wus/pWUS::WUS-linker-GFP/pCLV3::AlcR/pAlcA::NSlmb-vhhGFP4* plants to inducibly degrade WUS protein in stem cells (*9*). Ethanol induction reduced the WUS-linker–green fluorescent protein (GFP) signal in the SAM. In such plants, CMV invaded the SAM CZ and PZ (Fig. 2, M to T, and fig. S6). Therefore, the stem cell regulator WUS functions in inhibition of CMV infection and supports antiviral immunity in the SAM.

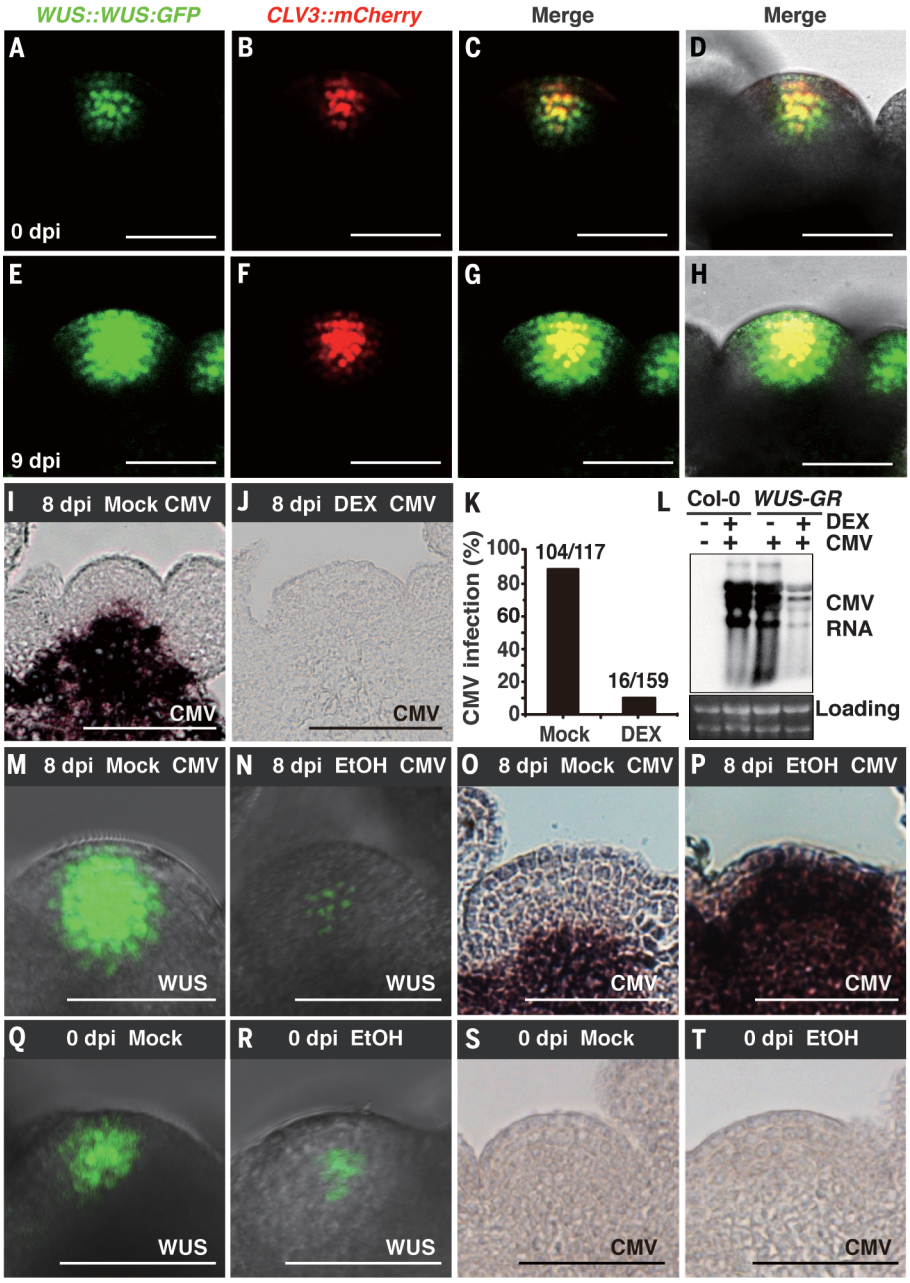


Fig. 2. WUS protein responds to CMV and inhibits viral infection. (**A** to **H**) WUS protein distribution and *CLV3* expression patterns in the *wus/pWUS::WUS-GFP; pCLV3::mCherry-NLS* rescue plant with [(E) to (H)] or without [(A) to (D)] CMV infection at 9 dpi (n = 15). (**I** and **J**) CMV accumulations at 8 dpi in *pUBQ10::WUS-GR* plants with (J) or without (I) DEX induction for 7 days. (**K**) Percentage of plants infected by viruses with (n = 159) or without (n = 117) DEX treatment at 8 dpi. (**L**) WUS inhibited CMV genomic RNA accumulation in the SAM, as determined by RNA gel blot at 8 dpi. (**M** to **T**) WUS-linker–GFP signals and CMV distributions in *wus/pWUS::WUS-linker-GFP/pCLV3::AlcR/pAlcA::NSlmb-vhhGFP4* lines with or without 1% ethanol treatment for 48 hours with CMV infection at 0 dpi [(Q) to (T)] and 8 dpi [(M) to (P)]. Scale bars, 50 μm in (A) to (H) and (M) to (T) and 100 μm in (I) and (J).

The role of WUS in antiviral immunity did not depend on RNA silencing (figs. S7 to S9), autophagy (fig. S10), or phytohormone (fig. S11) pathways, but WUS did lead to protein silencing in tobacco leaves (fig. S7). To determine whether WUS acts as a transcription factor in that context, we coexpressed *35S::WUS-GR* and *35S::GFP* in tobacco leaves and observed that GFP was silenced only if the translocation of WUS was induced (fig. S12A). We also found that the efficiency of transgene silencing was dependent on the dosage of *WUS-GR* (fig. S12B). Mutations in the WUS-box (m1) and EAR-like motif (m3), but not the acidic region motif (m2), abolished WUS functions in GFP silencing (fig. S13, A to F, I, and J). Moreover, we fused the C terminus of WUS-m1 with the SRDX repression domain or the VP16 activation domain and observed recovery of GFP silencing in *WUS-m1-SRDX*, but not in *WUS-m1-VP16* (fig. S13, G to J), suggesting that WUS acts as a transcriptional repressor to inhibit GFP accumulation.

To investigate whether WUS inhibits protein accumulation by accelerating protein degradation or by repressing protein synthesis, we used the proteasome inhibitor MG132. GFP degradation was repressed by MG132 treatment in tobacco leaves, allowing GFP accumulation. However, with the addition of WUS, GFP did not accumulate (fig. S14A). Similarly, in *Arabidopsis*, MG132 treatment did not affect WUS-mediated GFP degradation in *pUBQ10::WUS-GR*; *35S::GFP* plants (fig. S14, B and C). Thus, WUS inhibits protein synthesis rather than promoting protein degradation.

To examine the effect of WUS on viral protein synthesis, we fused GFP to the C terminus of each of the CMV proteins 2a, 3a, CP, and 2b (*10*). We coinfiltrated these constructs with *35S::WUS* in tobacco leaves. Expression of 3a-GFP, 2a-GFP, CP-GFP, and 2b-GFP was reduced by coinfiltration with *35S::WUS* [as shown by fluorescence and Western blot analysis (Fig. 3, A to C)]. To determine whether WUS represses CMV protein accumulation in the *Arabidopsis* SAM, we inoculated *pUBQ10::WUS-GR* plants with CMV and then induced WUS with DEX treatment. At 8 dpi, 2b protein abundance in the SAM decreased (Fig. 3D). To elucidate whether WUS-mediated protein inhibition is virus specific or if it has global effects on protein synthesis, we followed incorporation of *O*-propargyl-puromycin to assess nascent protein levels in *pUBQ10::WUS-GR* plants (*11*, *12*). Three days after DEX induction, we observed reduced total nascent protein content in *pUBQ10::WUS-GR* plants (Fig. 3, E and F), but not control plants (fig. S15), consistent with the cycloheximide treatment (Fig. 3, E and F). Thus, in both tobacco and *Arabidopsis*, WUS represses CMV protein accumulation and inhibits global protein synthesis.

To study how WUS inhibits protein synthesis, we profiled gene expression in apices of

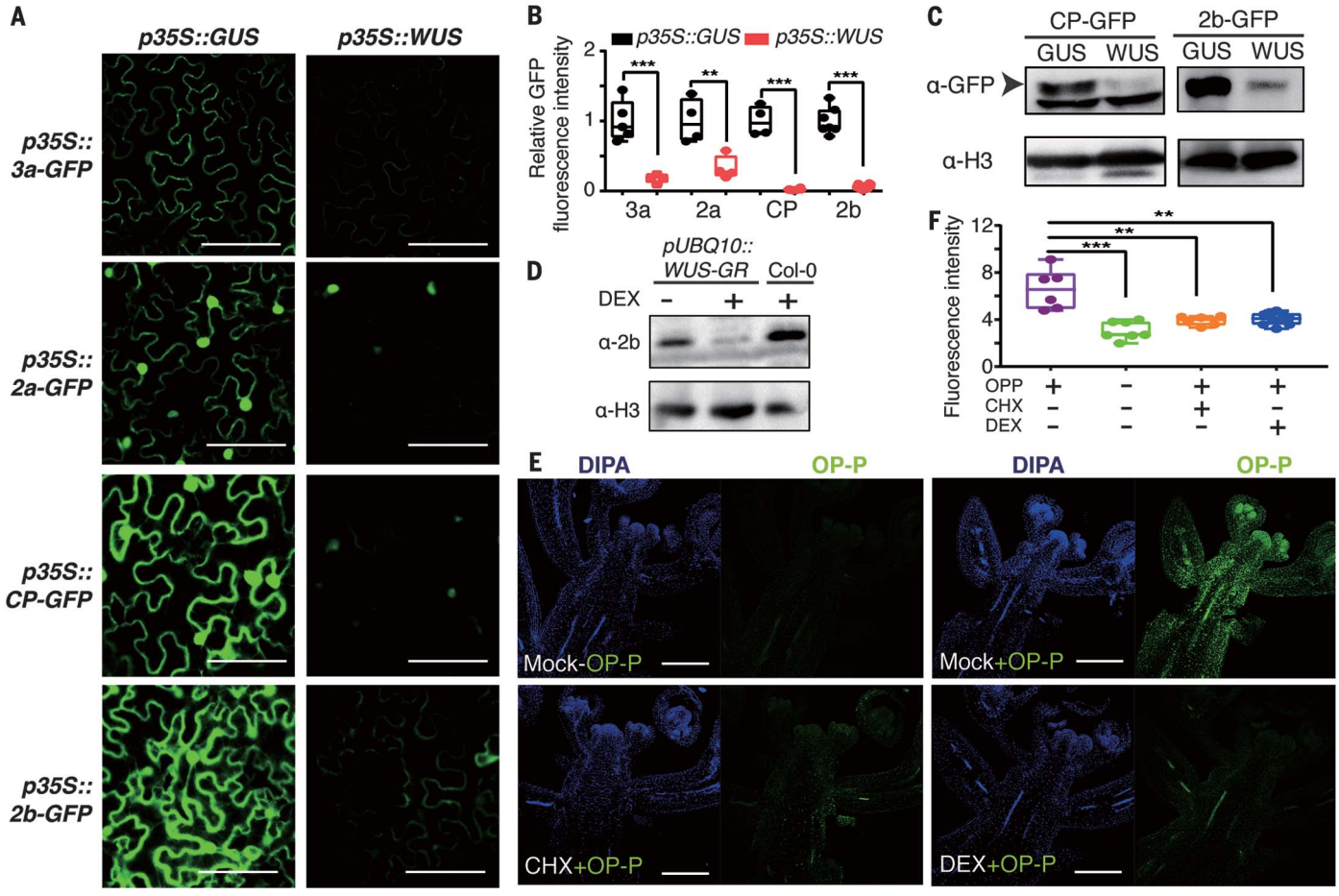


Fig. 3. WUS represses CMV protein accumulation in plants. (**A**) The CMV proteins 3a (*35S::3a-GFP*), 2a (*35S::2a-GFP*), CP (*35S::CP-GFP*), and 2b (*35S::2b-GFP*) were repressed in tobacco leaves 4 days after coinfiltration of *35S::WUS*. (**B**) Quantification of the relative fluorescence intensity of GFP in (A) ($n = 5$). (**C**) Western blots of CP and 2b proteins in (A). (**D**) WUS represses CMV 2b protein accumulation in the SAM of *Arabidopsis* at 8 dpi. (**E** and **F**) WUS globally inhibits nascent protein synthesis in *Arabidopsis* (E), and the relative fluorescence intensity is quantified in (F). ($n > 7$). *35S::GUS* was used as a negative control. CHX, cycloheximide; DIPA, 4′,6-diamidino-2-phenylindole; OP-P, *O*-propargyl-puromycin. Scale bars, 50 μm in (A) and 100 μm in (E). $^{**}P < 0.01$, $^{***}P < 0.001$, unpaired, two-tailed Student's *t* test.

DEX-induced *pUBQ10::WUS-GR* plants with or without CMV infection. We identified 11 genes involved in the regulation of protein synthesis, including 7 *S*-adenosyl-L-methionine–dependent methyltransferases (MTases) that are predicted to regulate RNA methylation (fig. S16). We examined these (fig. S17A) and 12 other known MTases (fig. S17B) using quantitative reverse transcription polymerase chain reaction (qRT-PCR). Six of the genes were repressed by WUS and up-regulated in the *wus-6* mutant (fig. S17, A to C). Five of these MTases contained a WUS DNA-binding motif. Indeed, we found with chromatin immunoprecipitation and electrophoretic mobility shift assays that WUS associates with the MTase promoter both in vivo and in vitro (fig. S17, D to I), suggesting that it represses the transcription of MTases in the SAM.

Among these five MTase genes, only *NUCLEOLAR PROTEIN 2A* (*NOP2A*, *AT5G55920*) showed expression in the PZ (fig. S18), and its expression was repressed by CMV infection (fig. S19). Because NOP2A and its homolog, NOP2B, were predicted to be MTases responsible for m^5C methylation at C2860 in 25*S* ribosomal RNA (rRNA) (*13*–*15*), we assessed total m^5C levels. We observed decreases in *nop2a/NOP2B* RNA interference (RNAi) transgenic plants (Fig. 4A), which we confirmed with liquid chromatography–tandem mass spectrometry (LC-MS/MS) quantification (Fig. 4B). RNA immunoprecipitation showed m^5C-modified 25*S* rRNAs enriched around C2860 in the wild-type plants and less so in *nop2a/NOP2B* RNAi plants (Fig. 4C), demonstrating that *NOP2A* and *NOP2B* act redundantly in m^5C methylation of 25*S* rRNAs in vivo. Using RNA bisulfite sequencing, we quantified the m^5C levels of 25*S* rRNA and observed a 64% increase of unmethylated C2860 in *nop2a/NOP2B* RNAi plants (Fig. 4D).

rRNA methylation stabilizes rRNA processing and ribosome structure (*16*). In MTase mutants, we observed disturbed pre-rRNA processing and increased rRNA intermediates (fig. S20), consistent with observations of WUS-overaccumulating plants (fig. S21). With polysome profiling, we observed reduced levels of 80*S* and 60*S* ribosomes in plants after activating WUS (Fig. 4E and fig. S22). Because *NOP2A* was repressed by WUS in stem cells, we observed less nascent protein synthesis in stem cells in an area coincident with the WUS protein domain (Fig. 4F). Although the catalytic activities of AT3G54150, AT1G69523, and AT1g55450 remain unclear, 80*S* and 60*S* ribosomes levels were also reduced in these three mutants (fig. S23). Thus, WUS represses ribosome assembly at least in part by negatively regulating MTases, resulting in a global decrease in protein synthesis.

Given that five MTase genes are directly repressed by WUS, we analyzed MTases in the context of a WUS-mediated antiviral mechanism. We coexpressed *p35S::WUS-GR* with MTase genes in tobacco leaves and observed that WUS-mediated GFP silencing was

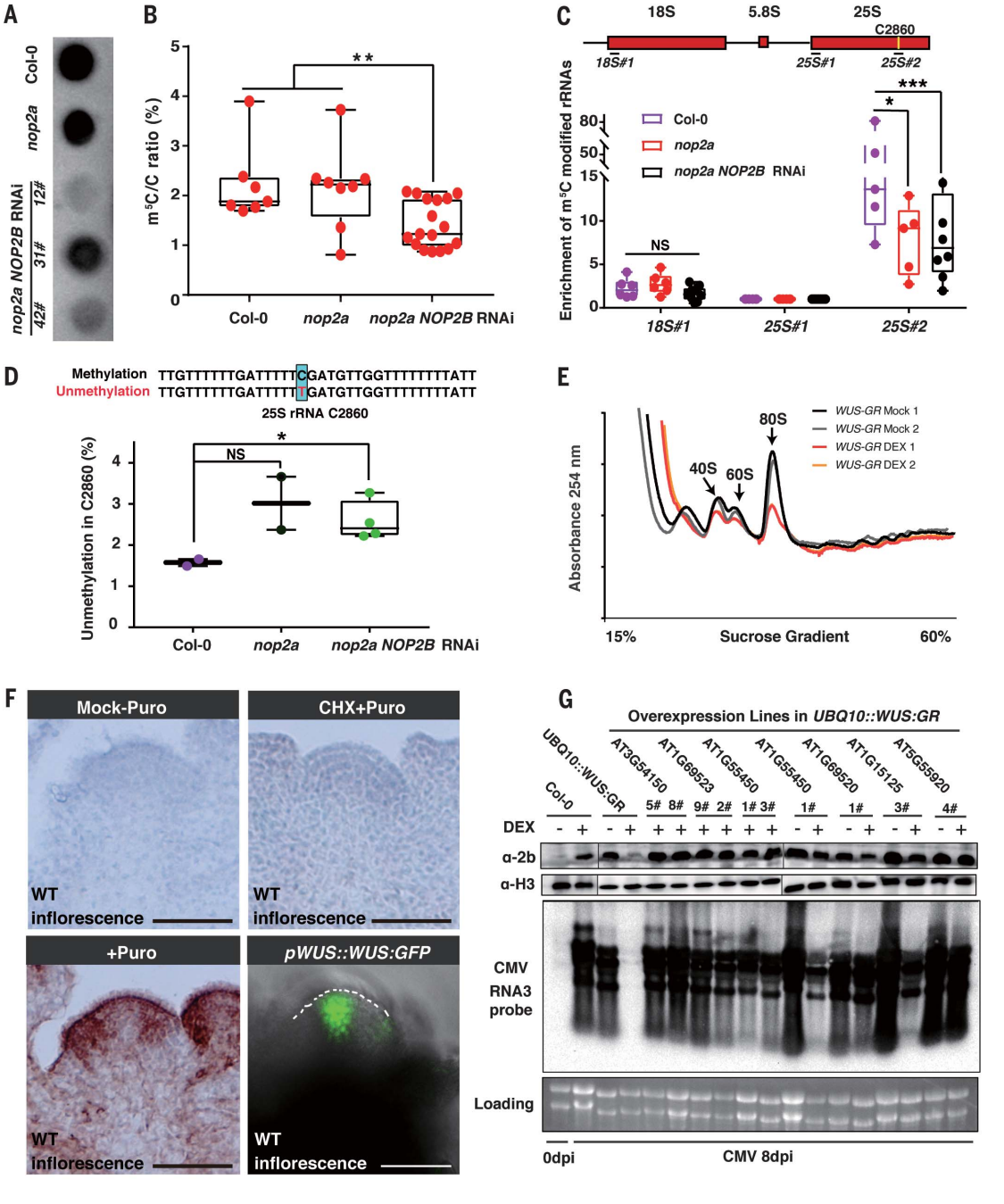


Fig. 4. MTase mediates WUS-triggered innate antiviral immunity. (**A**) Dot-blot assay showing that total m^5C levels were decreased in *nop2a/NOP2B* RNAi plants. (**B**) LC–MS/MS quantifications of the total m^5C levels of wild-type (WT) (n = 7), *nop2a* (n = 8), and independent *nop2a/NOP2B* RNAi T1 (n = 17) plants. (**C**) m^5C-RNA immunoprecipitation qRT-PCR of 4-week-old WT (n = 6), *nop2a* (n = 6), and independent *nop2a/NOP2B* RNAi T1 plants (n = 8). The region of 25*S*#2 contains the predicted m^5C target C2860 of 25*S* rRNA. The regions of 25*S*#1 and 18*S*#1 were used as the negative controls. (**D**) Quantification of the 25*S* rRNA m^5C levels of C2860 in WT (n = 2), *nop2a* (n = 2), and *nop2a/NOP2B* RNAi (n = 4) plants using RNA bisulfite sequencing. Black box indicates the C2860 site. (**E**) Polysome profile of 10-day-old *pUBQ10::WUS-GR* seedlings treated with DEX for 2 days (n = 2). (**F**) Nascent protein synthesis in the SAM. Inflorescences without puromycin (Puro) incorporation or with Puro and CHX treatment were used as the negative controls. Puro-incorporated nascent proteins were mainly observed in the PZ and primordia in areas complementary to the WUS protein domain. Scale bars, 50 μm. (**G**) Overexpression of MTases blocks WUS-mediated antiviral immunity. The shoot apices of MTases overexpression lines in the *UBQ10::WUS-GR* background with CMV infection at 8 dpi and DEX induction for 7 days were used for detection of the 2b protein with an anti-2b antibody and of the viral genomic RNA with the CMV RNA3 probe. *P < 0.05, **P < 0.01, ***P < 0.001, Student's t test. NS, no significant difference.

partially blocked by overexpression of MTases (fig. S24). We tested the genetic interaction of WUS and MTases in *Arabidopsis* during viral infection by overexpressing six MTases in *pUBQ10::WUS-GR* plants and observed increases in viral protein and genomic RNA accumulation upon WUS induction (Fig. 4G). Thus, WUS-mediated transcriptional repression of MTases is essential for innate antiviral immunity in plants.

Meristem tip culture has been used to produce virus-free plants of many species against a range of viral infections, which suggests a conserved and broad-spectrum antiviral mechanism in the SAM. We therefore tested three additional viruses that trigger systemic infection in *Arabidopsis*, including turnip crinkle virus (TCV), tobacco rattle virus (TRV), and turnip mosaic virus (TuMV) (fig. S25). We observed that WUS repressed multiple viral infections in the SAM of *Arabidopsis* (fig. S26). Given that WUSCHEL-related homeobox (WOX) genes are widespread in the plant kingdom and have conserved functions in shoot and root meristems (*17*, *18*), WUS/WOX–mediated antiviral immunity might be a conserved mechanism in plants.

As a parasitic strategy, viruses hijack the host protein synthesis machinery to complete their life cycle. Plants in turn have evolved various mechanisms to protect them against viral infection. Here, we show how WUS-triggered antiviral immunity safeguards stem cells and their nascent daughter cells from

viral invasion. In stem cells, WUS directly represses MTase transcription, impairing global protein synthesis and limiting the replication and spread of the virus (fig. S27).

REFERENCES AND NOTES

1. E. Aichinger, N. Kornet, T. Friedrich, T. Laux, *Annu. Rev. Plant Biol.* **63**, 615–636 (2012).
2. G. Morel, *CR Acad. Sci. Paris* **235**, 1324–1325 (1952).
3. P. J. Wang, C. Y. Hu, "Regeneration of virus-free plants through in vitro culture," in *Advances in Biomedical Engineering*, A. Fiechter, Ed. (Springer, 1980), vol. 18, pp. 61–99.
4. R. K. Yadav *et al*., *Genes Dev.* **25**, 2025–2030 (2011).
5. K. F. X. Mayer *et al*., *Cell* **95**, 805–815 (1998).
6. J. C. Fletcher, U. Brand, M. P. Running, R. Simon, E. M. Meyerowitz, *Science* **283**, 1911–1914 (1999).
7. U. Brand, J. C. Fletcher, M. Hobe, E. M. Meyerowitz, R. Simon, *Science* **289**, 617–619 (2000).
8. H. Schoof *et al*., *Cell* **100**, 635–644 (2000).
9. Y. Ma *et al*., *Nat. Commun.* **10**, 5093 (2019).
10. S.-W. Ding, B. J. Anderson, H. R. Haase, R. H. Symons, *Virology* **198**, 593–601 (1994).
11. S. Blanco *et al*., *Nature* **534**, 335–340 (2016).
12. J. Liu, Y. Xu, D. Stoleru, A. Salic, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 413–418 (2012).
13. U. Fujikura, G. Horiguchi, M. R. Ponce, J. L. Micol, H. Tsukaya, *Plant J.* **59**, 499–508 (2009).
14. A. L. Burgess, R. David, I. R. Searle, *BMC Plant Biol.* **15**, 199 (2015).
15. A. Pavlopoulou, S. Kossida, *Genomics* **93**, 350–357 (2009).
16. A. Gigova, S. Duggimpudi, T. Pollex, M. Schaefer, M. Koš, *RNA* **20**, 1632–1644 (2014).
17. K. Mukherjee, L. Brocchieri, T. R. Bürglin, *Mol. Biol. Evol.* **26**, 2775–2794 (2009).
18. J. Nardmann, P. Reisewitz, W. Werr, *Mol. Biol. Evol.* **26**, 1745–1755 (2009).

ACKNOWLEDGMENTS

We thank J. Wang, X. Zhang, F. Qu, X. Wang, Y. Xiong, C. Xiang, I.R. Searle, M. Sun, T. Laux, D. Yu, X. Yuan, J. Gao, and Y. Liu for sharing plant and virus materials. **Funding:** This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27030105 to Z.Z.), the National Natural Science Foundation of China (31870264 and 31570273 to Z.Z., 31400215 to H.W., 31871289 to Z.D.), the Ministry of Science and Technology of China (2013CB967300 to Z.Z.), the China Postdoctoral Science Foundation (2015M582004 to H.W.), and the Deutsche Forschungsgemeinschaft (SFB1101 to J.U.L).
Author contributions: Z.Z., H.W., and Z.T. designed the experiments, analyzed the data, and wrote the paper. H.W. performed ChIP, RNA in situ hybridization, RNA blot, polysome profiling, nascent protein assay, and LC-MS/MS. X.Q. and M.S. performed the Western blot and qRT-PCRs. Z.D., X.L., L.Z., Z.T., and Z.J. conducted RNA-seq and small RNA-seq and analyzed the obtained data. L.L., C.S., and M.X. performed the EMSA experiments. J.F. and J.U.L. generated the *pWUS::WUS-GFP; pCLV3::mCherry-NLS* and *pUBQ10::WUS-GR* transgenic lines. S.L. performed vector construction and plant transformation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw sequence data generated during this study are available in the Gene Expression Omnibus (GEO) repository under accession no. GSE128778 for RNA-seq and no. GSE128863 for small RNA-seq. Requests for materials should be addressed to Z.Z. All data presented in this study are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/227/suppl/DC1
Materials and Methods
Figs. S1 to S27
Table S1
References (*19*–*45*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

15 March 2020; resubmitted 28 July 2020
Accepted 21 August 2020
10.1126/science.abb7360

SUPERCONDUCTIVITY

Clean 2D superconductivity in a bulk van der Waals superlattice

A. Devarakonda[1], H. Inoue[1]*, S. Fang[2]†, C. Ozsoy-Keskinbora[3]‡, T. Suzuki[1], M. Kriener[4], L. Fu[1], E. Kaxiras[2,3], D. C. Bell[3,5], J. G. Checkelsky[1]§

Advances in low-dimensional superconductivity are often realized through improvements in material quality. Apart from a small group of organic materials, there is a near absence of clean-limit two-dimensional (2D) superconductors, which presents an impediment to the pursuit of numerous long-standing predictions for exotic superconductivity with fragile pairing symmetries. We developed a bulk superlattice consisting of the transition metal dichalcogenide (TMD) superconductor 2*H*-niobium disulfide (2*H*-NbS_2) and a commensurate block layer that yields enhanced two-dimensionality, high electronic quality, and clean-limit inorganic 2D superconductivity. The structure of this material may naturally be extended to generate a distinct family of 2D superconductors, topological insulators, and excitonic systems based on TMDs with improved material properties.

Superconducting states with real- or momentum-space nodes in the gap function are acutely sensitive to disorder. Such nodal pairing states are weakened by the momentum-space averaging of the gap function caused by disorder-induced scattering of Cooper pairs across the Fermi surface. Conventionally, this disorder averaging can be avoided when Cooper pairs have well-defined crystal momentum k within the Pippard coherence length ξ_0; this regime is realized when ξ_0 is smaller than the electronic mean free path ℓ (i.e., $\xi_0/\ell \ll 1$). Two-dimensional superconductors in this so-called clean limit play a central role in proposals including archetypal platforms for finite-momentum Cooper pairing (*1*, *2*) and recent constructions for unconventional superconducting phases that leverage normal-state spin textures (*3*, *4*). At present, there is a paucity of materials that allow access to this regime.

Figure 1A presents a historical survey of superconducting materials sorted according to their dimensionality and cleanliness, the former characterized by H_{c2}^{c}/H_{c2}^{ab} (the ratio of the upper critical magnetic fields perpendicular and parallel to the 2D layer) and the latter by ξ_0/ℓ. Early experimental work in granular Al (*g*-Al) and amorphous Bi (*a*-Bi) films demonstrated 2D superconductivity through precise control of the superconducting layer thickness; these systems were later used in groundbreaking studies that observed the vortex-antivortex unbinding transition of Berezinskii-Kosterlitz-Thouless (BKT) that had been predicted for 2D superconductors (*5*). Shortly thereafter, studies of similar films revealed that the disorder induced a superconductor-insulator transition in two dimensions (*6*). Further developments in thin-film growth have led to the observation of the BKT transition in crystalline films of Nb (*c*-Nb) (*7*) and, more recently, to high-quality 2D superconductors formed at epitaxial interfaces in $LaAlO_3/SrTiO_3$ (*8*) and in δ-doped Nb-$SrTiO_3$ (*9*).

In parallel, the study of anisotropic bulk superconductors has also yielded substantial insight into the nature of the superconducting state. These materials can also be framed in the context of 2D superconductivity (see Fig. 1A); prototypical examples include cuprates, pnictides, organic charge-transfer salts, graphite intercalation compounds, and transition metal dichalcogenides (TMDs). Considerable effort has been invested in pursuing tunable dimensionality in these systems. For example, the search for new cuprates in the Ruddlesden-Popper series and their oxygen-deficient variants (*10*) has led to a diverse set of materials with variable dimensionality and a rich variety of exotic behavior (*11*). Discovery of the structurally related layered perovskite Sr_2RuO_4 realized a 2D stoichiometric superconductor well within the clean limit, in principle satisfying the stability requirements for fragile superconducting gap functions that would otherwise be disrupted by nonmagnetic disorder (*12*). Despite a great deal of interest in such phases, this very clean and highly 2D regime has largely been limited to organic systems (*13*).

Substantial efforts have gone toward achieving 2D superconductivity in TMDs, which, given their intrinsically strong spin-orbit coupling and inversion symmetry breaking, are expected to yield exotic forms of superconductivity in the clean limit (*14*). As shown in Fig. 1A for the case of 2*H*-TaS_2, intercalation of organic molecules (*15*) [$(Py)_{0.5}$-TaS_2, where Py is pyridine] and incommensurate spacer layers (*16*) [$(PbS)_{1.13}TaS_2$] has been used to drive these materials toward the 2D limit and reduce the inversion-symmetric coupling between adjacent layers (*17*). These have led to important advances, such as establishing that superconductivity is a property of the 2D planes in $(Py)_{0.5}$-TaS_2. More recently, developments in the exfoliation of van der Waals (vdW) layered materials have made atomically thin 2D superconductors more readily accessible (e.g., *ML*-TaS_2) (*18*–*21*). Subsequent studies have revealed a form of superconductivity characterized by strong Ising spin-orbit coupling equivalent to applying magnetic fields on the order of 100 T (*19*–*21*). However, flakes exfoliated from the bulk are often subject to degradation and reduction in quality during the fabrication process (*22*). Here, we show that high-quality *H*-NbS_2 monolayers with electronic mobilities more than three orders of magnitude larger than in bulk 2*H*-NbS_2 can be realized in a bulk single-crystal superlattice formed with a commensurate block layer. Correspondingly, we find that this material is a clean-limit 2D superconductor exhibiting a BKT transition at T_{BKT} = 0.82 K and prominent 2D Shubnikov-de Haas (SdH) quantum oscillations.

The fundamental structural unit in hexagonal TMDs is the *H*-MX_2 layer, where *M* and *X* are a transition metal and chalcogen, respectively. As shown in Fig. 1B, this structure [point group symmetry $\bar{6}m2$ (D_{3h})] breaks inversion symmetry in the layer plane owing to the trigonal prismatic coordination of *X* around *M* (the missing inversion partners are shown as dashed circles), and as a result yields an out-of-plane (Ising) spin texture (*18*–*21*). For thin flakes deposited on substrates, the substrate-flake interface breaks mirror symmetry (Fig. 1C) and yields an in-plane (Rashba) spin texture (*23*). Taken together, for TMD flakes on substrates, the simultaneous breaking of mirror and inversion symmetry leads to a mixed spin texture (Fig. 1D) on the Fermi surface composed of both Ising and Rashba components (*18*). These spin textures and the resulting physics are suppressed in the bulk limit, where the overall unit cell preserves inversion symmetry.

We have synthesized a single-crystal material, $Ba_6Nb_{11}S_{28}$, composed of high-quality *H*-NbS_2 layers and Ba_3NbS_5 block layers in which the TMD layers are strongly decoupled (*17*). Figure 1E shows a cross section of the structure imaged by high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) with the model structure superimposed. As determined by electron and powder x-ray diffraction, the unit cell (space group $P\bar{3}1c$ with a = 10.4 Å, c = 24.5 Å) is composed of two inversion-related *H*-NbS_2 layers across each of which mirror symmetry is broken by the

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [3]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [4]RIKEN Center for Emergent Matter Science (CEMS), Wako 351-0198, Japan. [5]Center for Nanoscale Systems, Harvard University, Cambridge, MA 02138, USA.
*Present address: Institute for Materials Research, Tohoku University, Sendai, Miyagi 980-8577, Japan.
†Present address: Department of Physics and Astronomy, Center for Materials Theory, Rutgers University, Piscataway, NJ 08854, USA.
‡Present address: Thermo Fisher Scientific, 5600KA Eindhoven, Netherlands.
§Corresponding author. Email: checkelsky@mit.edu

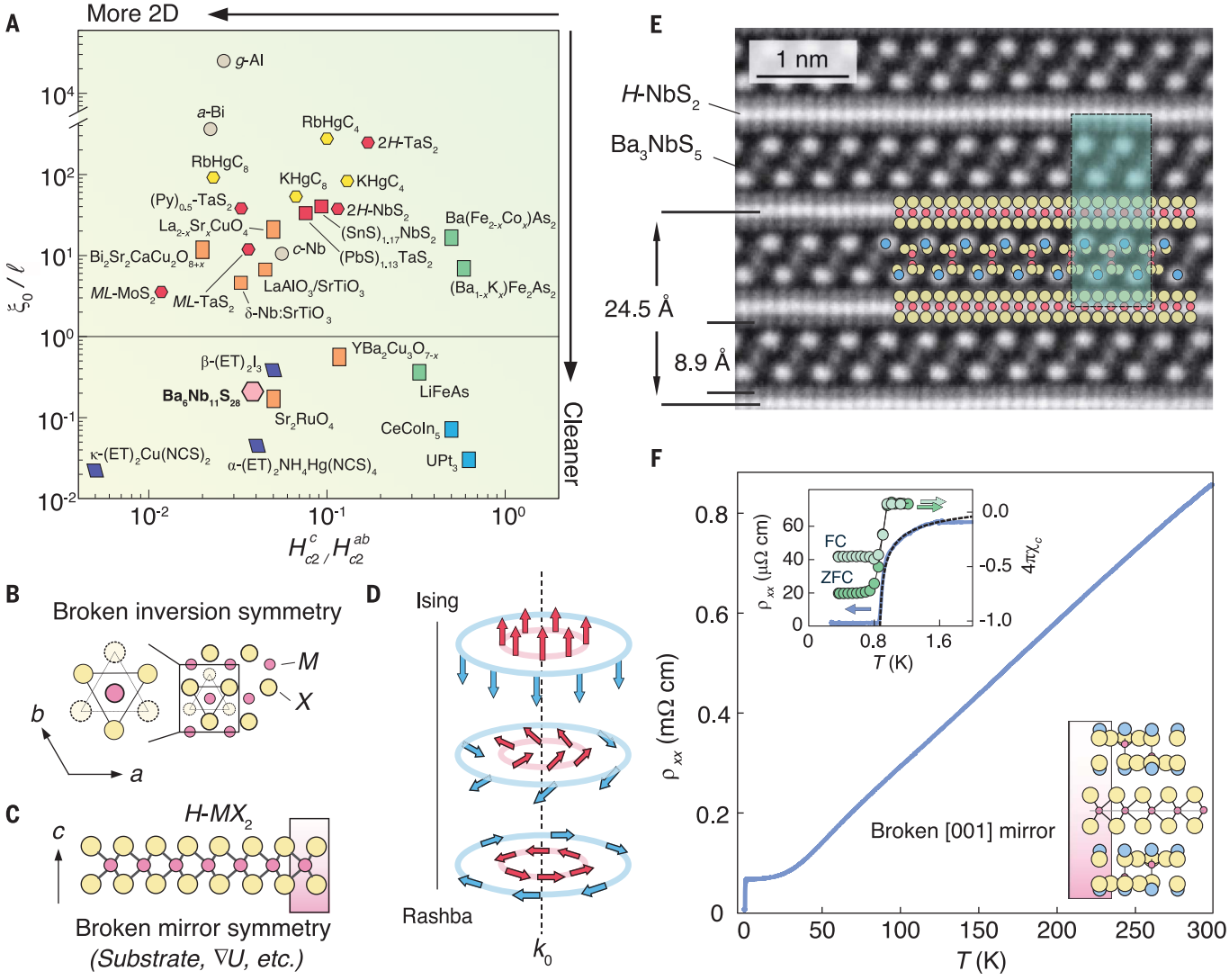


Fig. 1. 2D superconductivity and $Ba_6Nb_{11}S_{28}$. (**A**) Survey of superconducting materials characterized by anisotropy of the upper critical field H_{c2}^{c}/H_{c2}^{ab} and ratio of the Pippard coherence length to mean free path ξ_0/ℓ. The boundary between the clean and dirty limits is shown as a horizontal line. (**B**) Crystal structure of *H*-MX_2 projected onto the *ab*-plane. Lack of inversion symmetry is illustrated by the missing chalcogen (*X*) inversion partners (dashed circles). (**C**) The *ab*-plane mirror symmetry in monolayer *H*-MX_2 can be broken by substrates or local fields (∇U). (**D**) Depiction of momentum space spin-orbit texture for monolayer *H*-MX_2 with varying degrees of Ising and Rashba coupling. (**E**) HAADF-STEM image of $Ba_6Nb_{11}S_{28}$ taken along the $[1\bar{1}00]$ axis (scale bar, 1 nm). A simulation of the model structure is overlaid with one unit cell shaded in green. Ba, Nb, and S atoms are depicted as blue, red, and yellow circles, respectively. (**F**) Resistivity as a function of temperature $\rho_{xx}(T)$ in $Ba_6Nb_{11}S_{28}$ showing the superconducting transition. Upper inset: Magnified view of the transition in $\rho_{xx}(T)$ and magnetic susceptibility $4\pi\chi_c$ measured with zero field cooling (ZFC) and field cooling (FC). $\rho_{xx}(T)$ is well fit by the Halperin-Nelson model shown in black (see text). Lower inset: *H*-NbS_2 layer and mirror symmetry–breaking Ba_3NbS_5 block layers.

neighboring block layers; the overall unit cell retains inversion symmetry. The *H*-NbS_2 interlayer distance d = 8.9 Å is more than three times that of 2*H*-NbS_2 (*24*), leading to a reduction of the interlayer transfer integral $t_\perp$. This amplifies the two-dimensionality of the electronic structure relative to 2*H*-NbS_2 and enables local symmetry breaking–induced spin-orbit textures on the *H*-NbS_2 layers (*25*, *26*). [*H*-NbS_2 monolayers in $Ba_6Nb_{11}S_{28}$ experience local inversion symmetry breaking with point group symmetry 32 (D_3) (*17*).] Whereas traditional misfit compounds combine incommensurate layers in a superlattice, $Ba_6Nb_{11}S_{28}$ exhibits a 3 × 3 in-plane, commensurate superstructure caused by the difference in in-plane lattice constants between the two layer types (*17*), which leads to additional modification of the electronic structure.

Figure 1F shows the dependence of electrical resistivity $\rho_{xx}(T)$ on temperature T for $Ba_6Nb_{11}S_{28}$. The system is a metal, eventually showing superconductivity below T = 1 K. This can be compared to bulk 2*H*-NbS_2, which is also metallic and becomes a superconductor at critical temperature T_c = 5.7 K. Unlike several other related *H*-MX_2 systems, neither $Ba_6Nb_{11}S_{28}$ nor 2*H*-NbS_2 shows signs of a density wave transition (*17*, *27*). The upper inset of Fig. 1F shows a detailed view of the superconducting transition, which shows onset near T = 1.6 K and reaches zero resistance at T = 0.85 K. At the latter temperature, the magnetic susceptibility $4\pi\chi_c$ with field along the *c* axis shows a Meissner signal reaching a shielding fraction of 75% (Fig. 1F, upper inset, dark green ZFC data points) and a volume fraction of 40% (Fig. 1F, upper inset, light green FC data points). The reduction in the transition temperature for the NbS_2 layers relative to the bulk is similar to that observed in organic intercalated variants (*28*) and is consistent with

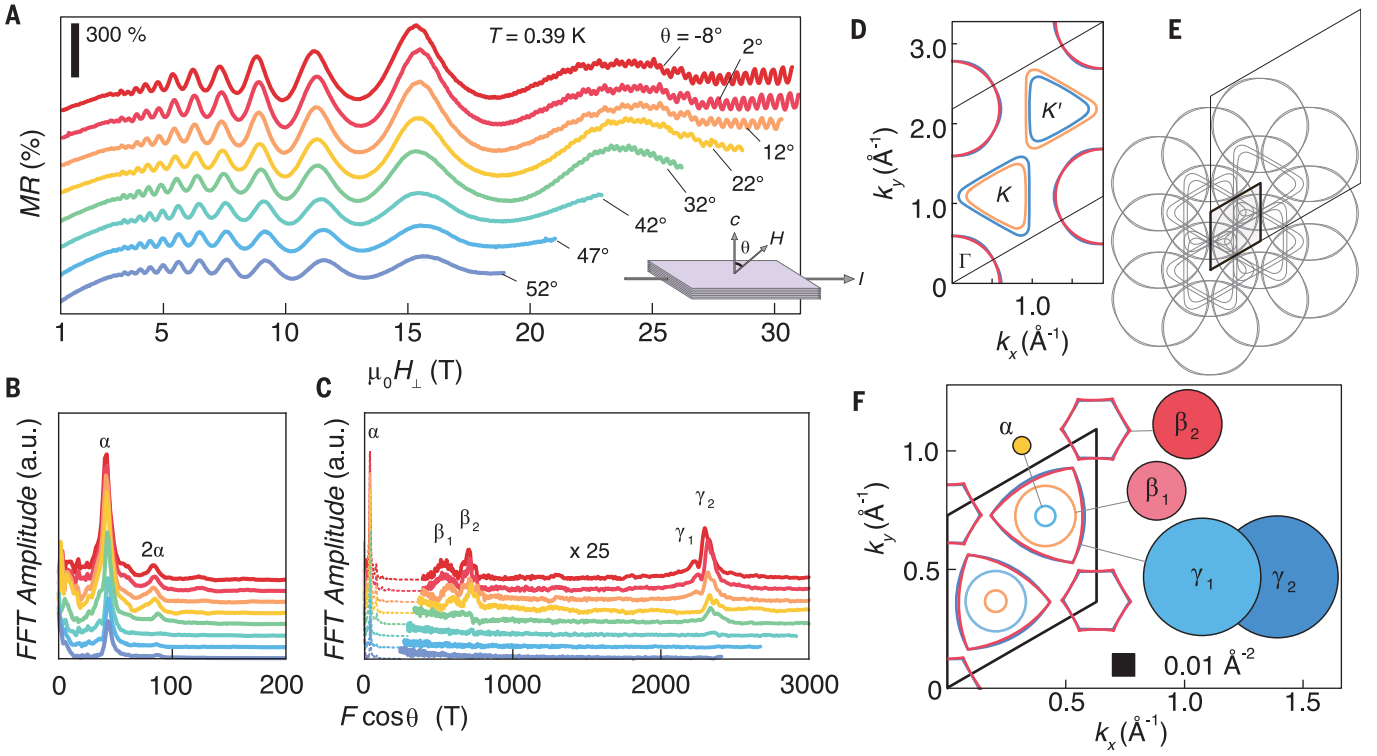


Fig. 2. Quantum oscillations and electronic structure of $Ba_6Nb_{11}S_{28}$. (**A**) Magnetoresistance as a function of perpendicular field $MR \equiv [\rho_{xx}(\mu_0 H_\perp)/\rho_{xx}(0)] - 1$ at temperature T = 0.39 K for different field rotation angles θ (geometry defined as shown in the inset). Curves are vertically offset by 150% of *MR* for clarity. (**B** and **C**) Low-frequency range (B) and full range (C) of quantum oscillation amplitude FFT as a function of perpendicular frequency $F\cos\theta$. The FFT amplitudes for the higher-frequency pockets are multiplied by 25. (**D**) DFT calculation of monolayer *H*-NbS_2 Fermi surfaces including spin-orbit coupling (*17*). (**E**) Depiction of zone-folding scheme involving the 3 × 3 superstructure imposed by the Ba_3NbS_5 block layer where the reduced Brillouin zone is enclosed by the bold line. (**F**) Electronic structure of zone-folded monolayer *H*-NbS_2 with experimentally observed Fermi surface cross-sectional areas drawn to scale as solid circles. The black box corresponds to 0.01 $Å^{-2}$.

a reduction of the electron-phonon coupling strength inferred from the measured resistivity at high T (*17*).

The temperature dependence of resistivity for 0.85 K < T < 1.6 K is well described by the Halperin-Nelson model, $\rho_{xx}^F(T) = \rho_{xx}^N \exp(-b/\sqrt{t})$, where ρ_{xx}^F and ρ_{xx}^N are the fluctuation and normal-state resistivity, respectively; $t = (T/T_{HN}) - 1$; and b is a fitting parameter on the order of 1 (Fig. 1F, upper inset, dashed curve) (*29*). The agreement with the Halperin-Nelson model evidences fluctuations of the superconducting order parameter above a two-dimensional BKT transition. Such behavior is generally rare in bulk single crystals but has been reported in $La_{1.875}Ba_{0.125}CuO_4$ and attributed to the decoupling of superconducting CuO_2 planes by stripe order (*30*). This is a further departure from the behavior in bulk 2*H*-NbS_2, which exhibits a sharp superconducting transition (*17*) (fig. S23); instead, it closely resembles those observed in monolayer *H*-MX_2. Further evidence of increased two-dimensionality is the resistivity anisotropy $\rho_{zz}/\rho_{xx} > 10^3$ at low T in the normal state of $Ba_6Nb_{11}S_{28}$ (where ρ_{zz} is the *c*-axis resistivity), which is substantially enhanced relative to $\rho_{zz}/\rho_{xx} \sim 100$ in 2*H*-NbS_2 (*17*) (fig. S24).

Magnetotransport measurements demonstrate the cleanliness of this material and show further evidence for a 2D electronic structure. Figure 2A shows the magnetoresistance $MR \equiv [\rho_{xx}(H)/\rho_{xx}(0)] - 1$, measured to 31 T. We observe SdH quantum oscillations that respond to the component of the magnetic field perpendicular to the *ab*-plane (the tilt angle θ is measured between the *c* axis and the applied field). The fast Fourier transform (FFT) computed after subtracting a monotonically increasing background (*17*) (fig. S14) plotted versus inverse field shows this more clearly (Fig. 2, B and C). Here, the oscillation frequency multiplied by cos(θ) has little variance versus angle, demonstrating the 2D nature of the Fermi surface. This is qualitatively different from 2*H*-NbS_2, for which electronic structure calculations indicate warped and elliptical Fermi surfaces (*31*). Owing to the reduced coupling between the TMD layers, the observed bands (labeled here as α, β_1, β_2, γ_1, and γ_2) can be understood by starting with the 2D electronic structure of monolayer *H*-NbS_2, which consists of bands at the Γ, *K*, and *K*′ points of the hexagonal Brillouin zone (Fig. 2D), and zone folding into the reduced Brillouin zone determined by the 3 × 3 superstructure imposed by the block layers (Fig. 2E). In particular, the reduction in the pocket size from monolayer *H*-NbS_2 caused by zone folding, approximately one order of magnitude, quantitatively captures the size of the observed pockets (Fig. 2F) and is further supported by first-principles calculations (*17*) (fig. S12). An important aspect of this structure is that the large ratio of the spin-orbit coupling to $t_\perp$, evident from the degree of two-dimensionality, enables local symmetry breaking to affect the bulk electronic structure. The zone folding promotes the Rashba-textured pockets associated with the Γ point in monolayer *H*-NbS_2 to be larger than the Ising-split pockets at *K* and *K*′ [supported by comparing the calculated band structure for monolayer *H*-NbS_2 with that of the 3 × 3 zone-folded structure (*17*) (figs. S11 and S12)] and has potential implications for superconducting pairing.

More generally, it is noteworthy that quantum oscillations have not been reported in 2*H*-NbS_2; there, the typical transport mobilities reported for bulk single crystals are on the

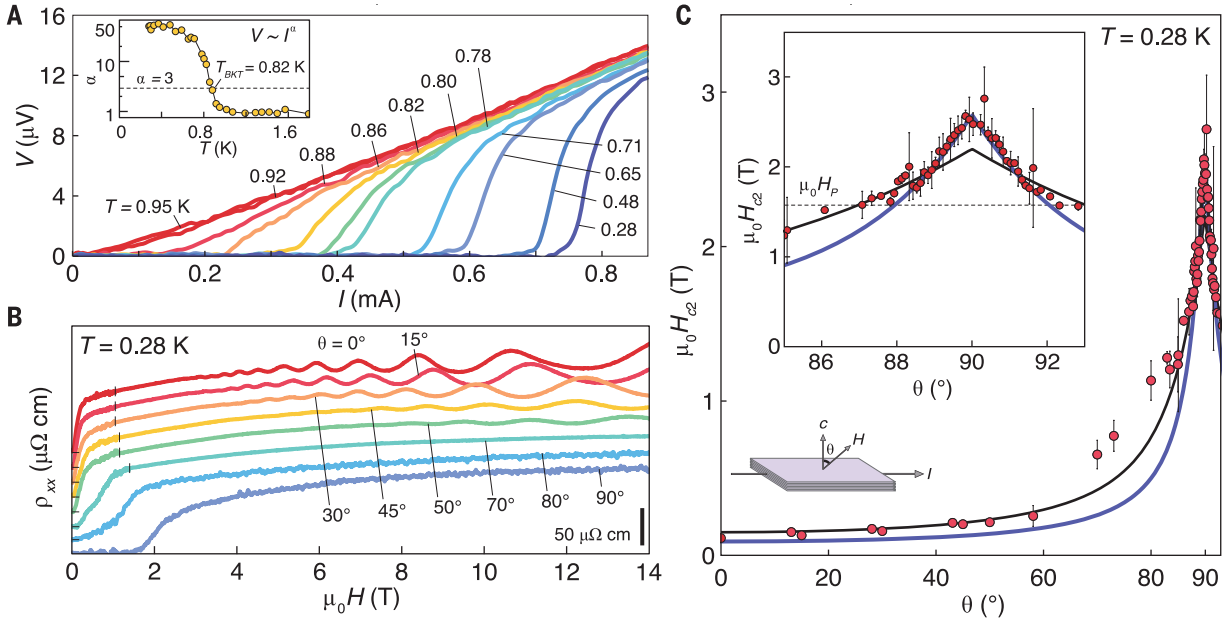


Fig. 3. 2D superconductivity and Pauli limit breaking in $Ba_6Nb_{11}S_{28}$. (**A**) Current-voltage characteristics $I(V)$ from T = 0.95 K to T = 0.28 K. The inset shows the evolution of the power law $V \propto I^{\alpha}$; the horizontal line marks $\alpha = 3$. (**B**) Longitudinal resistivity ρ_{xx} as a function of field $\mu_0 H$ for different values of θ. Curves are vertically offset by 20 μΩ·cm for clarity (horizontal lines). Vertical ticks separate regions measured with low current (7 μA) and higher current (70 μA) to avoid suppression of superconductivity by Joule heating. For θ = 80° and 90°, only low current is used. (**C**) Angular dependence of upper critical field $\mu_0 H_{c2}$ measured at T = 0.28 K with fits to the 2D-Tinkham model, computed using data in the range $|\theta - 90°| < 1.7°$ (purple curve) and $|\theta - 90°| > 1.7°$ (black curve), respectively. The inset shows a detailed view near θ = 90° where an enhancement of $\mu_0 H_{c2}(\theta)$ is observed across the Pauli limit $\mu_0 H_P$. We define $\mu_0 H_{c2}$ and its error to be when ρ_{xx} reaches 50 ± 5% of the normal-state value.

order of 1 $cm^2\,V^{-1}\,s^{-1}$ (*27*). In $Ba_6Nb_{11}S_{28}$, we see the onset of SdH quantum oscillations in magnetic fields between 2 and 3 T, indicating quantum mobilities on the order of 10^3 $cm^2\,V^{-1}\,s^{-1}$. Analysis of the quantum oscillations and low-field magnetoresistance indicates an associated transport mean free path ℓ = 1.21 μm, which greatly exceeds the Pippard coherence length $\xi_0 \approx 0.18\hbar v_F/k_B T_c$ = 254 nm (where $\hbar$ is the Planck constant divided by 2π, v_F is the Fermi velocity, and k_B is the Boltzmann constant) (*17*). This places $Ba_6Nb_{11}S_{28}$ in the clean limit of superconductivity (Fig. 1A).

Turning to properties of the superconducting state, Fig. 3A shows the current-voltage $I(V)$ characteristics of $Ba_6Nb_{11}S_{28}$ across the superconducting transition. As expected for a BKT transition (*29*), a linear response at T = 0.95 K and above crosses over to a nonlinear dependence $V \propto I^{\alpha}$ with $\alpha \sim 3$ at T_{BKT} = 0.82 K, consistent with T_{HN} = 0.85 K. With further examination of the fluctuation conductivity and the slope of the power-law exponent close to T_{BKT}, we find evidence for a vanishingly small interlayer coupling in the superconducting state (*17*, *32*). Figure 3B shows the evolution of $\rho_{xx}(H)$ as a function of magnetic field for different values of θ. Whereas for θ = 0° superconductivity is suppressed with relatively low fields and gives rise to quantum oscillations, for larger θ the upper critical field $\mu_0 H_{c2}$ rapidly increases (we define $\mu_0 H_{c2}$ and its error to be when ρ_{xx} reaches 50 ± 5% of the normal-state value). Figure 3C summarizes this behavior, with $\mu_0 H_{c2}(\theta)$ showing a sharp cusp for in-plane fields. Recent studies of 2D superconductors have shown that a feature that distinguishes such systems from anisotropic 3D superconductors is the profile of $\mu_0 H_{c2}(\theta)$ following the 2D Tinkham form

$$\left(\frac{H_{c2}(\theta)\sin\theta}{H_{c2}^{ab}}\right)^2 + \left|\frac{H_{c2}(\theta)\cos\theta}{H_{c2}^{c}}\right| = 1 \qquad (1)$$

where H_{c2}^{ab} and H_{c2}^{c} are the upper critical fields for field applied in-plane and out-of-plane, respectively (*33*). The response of $Ba_6Nb_{11}S_{28}$ can be fit by such a form, contrasting the anisotropic 3D character of 2*H*-NbS_2 (*17*) (fig. S25). Furthermore, for $Ba_6Nb_{11}S_{28}$, we observe an enhancement of the scale of $\mu_0 H_{c2}(\theta)$ for angles below 1.7° measured relative to the *ab*-plane. As shown in the inset of Fig. 3C, this anomalous enhancement coincides with $\mu_0 H_{c2}(\theta)$ crossing the Pauli paramagnetic limit $\mu_0 H_P \approx 1.84 T_{BKT}$ = 1.51 T. We find that two independent Tinkham fits trace the data across the entire angular regime: For $|\theta - 90°| \geq 1.7°$ we have $\mu_0 H_{c2}^{c}$ = 0.15 T, $\mu_0 H_{c2}^{ab}$ = 2.19 T, and for $|\theta - 90°| \leq 1.7°$ we obtain $\mu_0 H_{c2}^{c}$ = 0.09 T, $\mu_0 H_{c2}^{ab}$ = 2.55 T.

To further examine the anomaly in $\mu_0 H_{c2}$, we measured $\rho_{xx}(H, T)$ with θ systematically tuned away from 90°. Plotted as the excess conductivity $\delta\sigma \equiv 1 - (\rho_{xx}/\rho_{xx}^{N})$, the substantial enhancement at low T and high H quickly disappears as θ is moved away from 90°, and by θ = 86° there is little variation with further field tilt (Fig. 4A). A distinct feature at all values of θ is the finite δσ associated with fluctuating superconductivity for low H extending to T beyond T_{BKT}. To remove this fluctuation contribution, we plot the difference δσ(θ = 90°) − δσ(θ = 84°) in Fig. 4B. The expected 2D Ginzburg-Landau behavior is shown as a green curve; the transition curve follows this response below T_{BKT} until $T/T_{BKT} \approx 0.6$, below which a considerable enhancement is observed. As shown in Fig. 4C, this behavior is confined to low temperature and to a small angular region δθ about the *ab*-plane.

These observations taken together indicate the appearance of a clean 2D superconducting state with enhanced stability (larger H_{c2}) when $T < 0.5 T_c$ and field $H > H_P$ applied very close to the layer plane. Various theoretical scenarios have been discussed for Pauli breaking in 2D superconductors, including spin-orbit scattering (*34*), Ising superconductivity (*19*–*21*), and Fulde-Ferrell-Larkin-Ovchinnikov (FFLO) states (*1*, *2*). Given the clean-limit superconductivity realized here, spin-orbit scattering enhancements cannot account for the present observations. The dominant local Rashba spin-orbit coupling in the present system reduces the importance of the local Ising coupling (*35*), and in Ising superconductors no

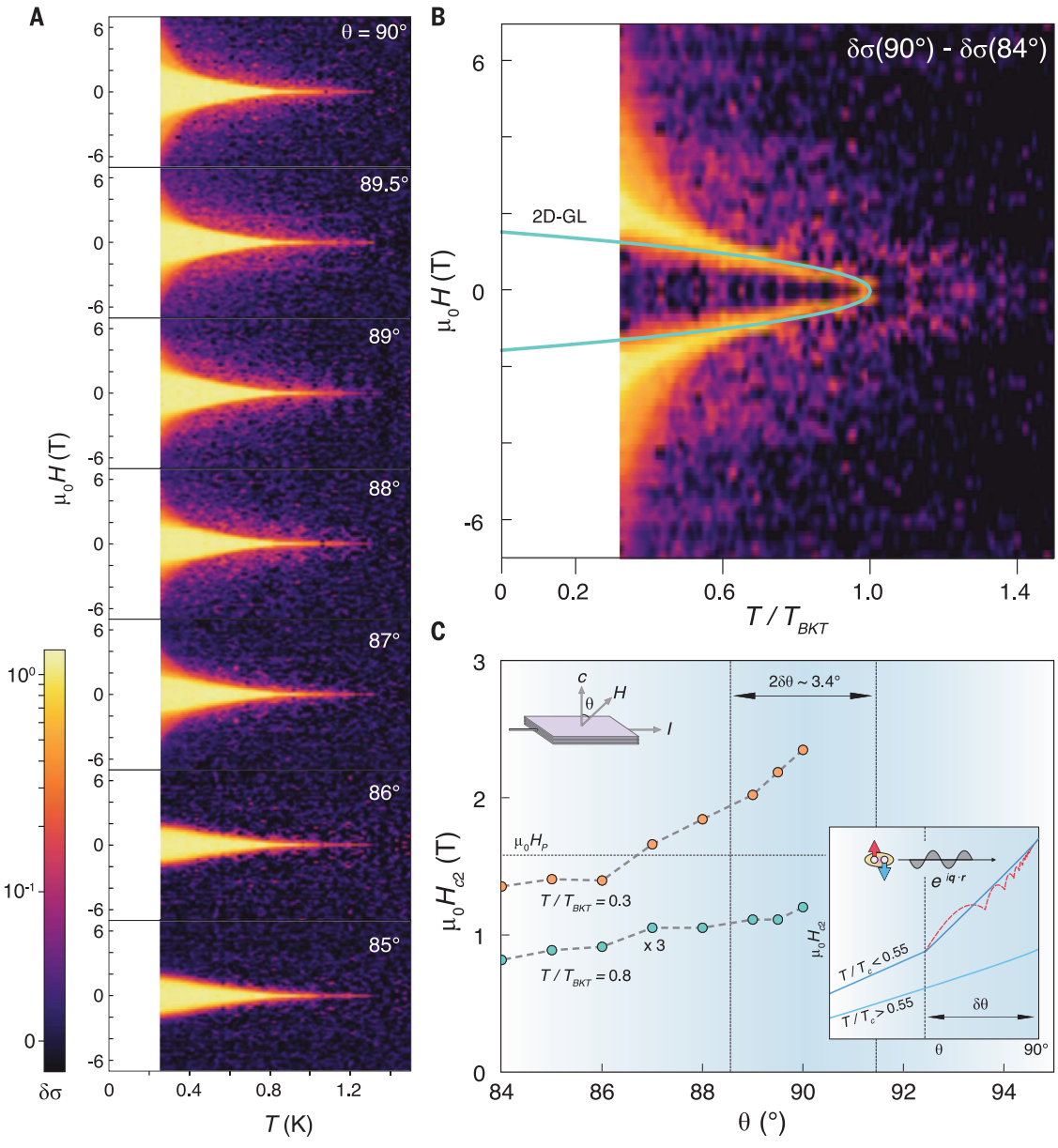


Fig. 4. Superconducting phase diagram of $Ba_6Nb_{11}S_{28}$. (**A**) Excess conductivity relative to the normal state $\delta\sigma(\mu_0H, T)$ for field angles θ near the *ab*-plane (θ = 90°). (**B**) Difference between $\delta\sigma(\mu_0H, T)$ for θ = 90° and θ = 84°. The temperature axis is normalized to T_{BKT}. The green curve represents the 2D Ginzburg-Landau (2D-GL) model of μ_0H_{c2}. (**C**) Angular dependence of μ_0H_{c2} at T/T_{BKT} = 0.3 (orange) and μ_0H_{c2} at T/T_{BKT} = 0.8 (green, magnified by a factor of 3). Inset: Schematic depiction of μ_0H_{c2} in a clean 2D system where an enhancement is expected within a critical region $|\theta - 90°| < \delta\theta$ where orbital limiting is quenched and an FFLO state, characterized by a superconducting gap Δ of the form $\exp(iq\cdot r)$, appears for $T/T_c < 0.55$ and $H > H_P$ (dark blue solid line). Theoretical studies of 2D FFLO superconductors further predict a cascade of magnetic vortex states that appear as a corrugation of $\mu_0H_{c2}(\theta)$ within this regime (*37*) (red dashed line).

abrupt change in $H_{c2}(\theta)$ is expected. Instead, the sharp angle–dependent enhancement resembles that of the clean layered organic FFLO candidate β″-$(ET)_2SF_5CH_2CF_2SO_3$ [where ET is bis(ethylenedithio)tetrathiafulvalene] in similar conditions of H and T arising when the orbital limiting of superconductivity is quenched by aligning H close to the layer plane (*36*). As shown schematically in the inset of Fig. 4C, once orbital limiting is overcome within a critical angular window δθ [typically on the order of 1° to 2° in organic systems (*37*)], one naturally expects a sharp enhancement of $H_{c2}(\theta)$ as the large in-plane magnetic field stabilizes a finite-momentum pairing state.

The close parallel between our observations and those expected for a 2D FFLO phase [along with our ability to fit the phase diagram in Fig. 4B with predictions for such a phase (*17*)] highlights it as a promising candidate. Theoretical studies of 2D FFLO superconductors further predict a cascade of magnetic vortex states with finite momentum for $T/T_c < 0.55$, $H > H_P$, and $|\theta - 90°| < \delta\theta$ (Fig. 4C, inset, red dashed line), which may be resolved by higher-resolution measurements. More directly, relative to clean 2D organic superconductors (Fig. 1A), the robust inorganic nature of $Ba_6Nb_{11}S_{28}$ offers the opportunity to examine the potential real-space modulation of superconductivity—using, for example, scattering techniques (*38*). Unlike other clean inorganic systems, $Ba_6Nb_{11}S_{28}$ is derived from a well-studied family of materials, lacks localized magnetic moments, and, perhaps most important, has vdW layer bonding that allows for

exfoliation and integration into superconducting device structures. One exciting consequence could be simplified fabrication of phase-sensitive junction devices similar to those used to study the cuprates and other unconventional superconductors (*39*).

We hypothesize that the large enhancement of electronic mobility observed for the *H*-NbS_2 layers in $Ba_6Nb_{11}S_{28}$ is attributable to screening by the highly polarizable block layer akin to that observed in engineered semiconductor heterostructures (*40*). Additionally, our density functional theory (DFT) calculations suggest that the lowest-energy cleavage occurs between the *H*-MX_2 and block layers, implying that mechanically exfoliated $Ba_6Nb_{11}S_{28}$ may yield naturally encapsulated *H*-NbS_2 monolayers akin to vdW structures made by stacking MX_2 layers and h-BN (*22*). [We observe, using optical and atomic force microscopy, that standard exfoliation techniques can be used to obtain flakes suitable for device fabrication (*17*).] However, given that bulk $Ba_6Nb_{11}S_{28}$ already exhibits two-dimensional physics, we propose that insertion of commensurate spacer layers could be an alternative to fabricating exfoliated nanodevices. There is also scope for functionalizing the spacer layer to further modulate the *H*-MX_2 layer—for example, by introducing magnetic constituents. The large electronic mean free path of $Ba_6Nb_{11}S_{28}$ enables clean-limit superconductivity and can potentially realize unconventional phases predicted in monolayer *H*-MX_2 superconductors (*3*, *4*, *41*, *42*). The physics of other MX_2 materials can also benefit from longer electron mean free paths. In particular, extending the materials family of MX_2 natural commensurate superlattices may, for example, pave the way to longer mean free paths, which would enable topological edge mode circuitry in WTe_2 superlattices (*43*) or longer exciton lifetimes in MoS_2 and related semiconducting TMD materials (*44*).

REFERENCES AND NOTES

1. L. N. Bulaevskii, *Sov. Phys. JETP* **38**, 634–639 (1973).
2. L. N. Bulaevskii, *Sov. Phys. JETP* **37**, 1133–1136 (1973).
3. B. T. Zhou, N. F. Q. Yuan, H. L. Jiang, K. T. Law, *Phys. Rev. B* **93**, 180501 (2016).
4. C.-X. Liu, *Phys. Rev. Lett.* **118**, 087001 (2017).
5. A. F. Hebard, A. T. Fiory, *Phys. Rev. Lett.* **44**, 291–294 (1980).
6. D. B. Haviland, Y. Liu, A. M. Goldman, *Phys. Rev. Lett.* **62**, 2180–2183 (1989).
7. J. W. P. Hsu, A. Kapitulnik, *Phys. Rev. B* **45**, 4819–4835 (1992).
8. N. Reyren *et al*., *Science* **317**, 1196–1199 (2007).
9. Y. Kozuka *et al*., *Nature* **462**, 487–490 (2009).
10. C. Park, R. L. Snyder, *J. Am. Ceram. Soc.* **78**, 3171–3194 (1995).
11. P. A. Lee, N. Nagaosa, X.-G. Wen, *Rev. Mod. Phys.* **78**, 17–85 (2006).
12. A. P. Mackenzie, Y. Maeno, *Rev. Mod. Phys.* **75**, 657–712 (2003).
13. J. Singleton, C. Mielke, *Contemp. Phys.* **43**, 63–96 (2002).
14. M. Smidman, M. B. Salamon, H. Q. Yuan, D. F. Agterberg, *Rep. Prog. Phys.* **80**, 036501 (2017).
15. F. R. Gamble, F. J. Disalvo, R. A. Klemm, T. H. Geballe, *Science* **168**, 568–570 (1970).
16. A. Meerschaut, *Curr. Opin. Solid State Mater. Sci.* **1**, 250–259 (1996).
17. See supplementary materials.
18. D. Xiao, G. B. Liu, W. Feng, X. Xu, W. Yao, *Phys. Rev. Lett.* **108**, 196802 (2012).
19. J. M. Lu *et al*., *Science* **350**, 1353–1357 (2015).
20. Y. Saito *et al*., *Nat. Phys.* **12**, 144–149 (2016).
21. X. Xi *et al*., *Nat. Phys.* **12**, 139–143 (2016).
22. S. Manzeli, D. Ovchinnikov, D. Pasquier, O. V. Yazyev, A. Kis, *Nat. Rev. Mater.* **2**, 17033 (2017).
23. Y. A. Bychkov, E. I. Rashba, *JETP Lett.* **39**, 78–81 (1984).
24. B. Morosin, *Acta Crystallogr. B* **30**, 551–552 (1974).
25. X. Zhang, Q. Liu, J. W. Luo, A. J. Freeman, A. Zunger, *Nat. Phys.* **10**, 387–393 (2014).
26. J. M. Riley *et al*., *Nat. Phys.* **10**, 835–839 (2014).
27. M. Naito, S. Tanaka, *J. Phys. Soc. Jpn.* **51**, 219–227 (1982).
28. Y. Hamaue, R. Aoki, *J. Phys. Soc. Jpn.* **55**, 1327–1335 (1986).
29. B. I. Halperin, D. R. Nelson, *J. Low Temp. Phys.* **36**, 599–616 (1979).
30. Q. Li, M. Hücker, G. D. Gu, A. M. Tsvelik, J. M. Tranquada, *Phys. Rev. Lett.* **99**, 067001 (2007).
31. C. Heil, M. Schlipf, F. Giustino, *Phys. Rev. B* **98**, 075120 (2018).
32. D. H. Kim, A. M. Goldman, J. H. Kang, R. T. Kampwirth, *Phys. Rev. B* **40**, 8834–8839 (1989).
33. M. Tinkham, *Introduction to Superconductivity* (Dover, ed. 2, 2004).
34. R. A. Klemm, A. Luther, M. R. Beasley, *Phys. Rev. B* **12**, 877–891 (1975).
35. J. Falson *et al*., *Science* **367**, 1454–1457 (2020).
36. R. Beyer, B. Bergk, S. Yasin, J. A. Schlueter, J. Wosnitza, *Phys. Rev. Lett.* **109**, 027003 (2012).
37. H. Shimahara, D. Rainer, *J. Phys. Soc. Jpn.* **66**, 3591–3599 (1997).
38. A. Huxley *et al*., *Nature* **406**, 160–164 (2000).
39. D. J. Van Harlingen, *Rev. Mod. Phys.* **67**, 515–535 (1995).
40. D. Jena, A. Konar, *Phys. Rev. Lett.* **98**, 136805 (2007).
41. Y. Nakamura, Y. Yanase, *Phys. Rev. B* **96**, 054501 (2017).
42. L. Wang, T. O. Rosdahl, D. Sticlet, *Phys. Rev. B* **98**, 205411 (2018).
43. Z. Fei *et al*., *Nat. Phys.* **13**, 677–682 (2017).
44. G. Wang *et al*., *Rev. Mod. Phys.* **90**, 021001 (2018).
45. A. Devarakonda, H. Inoue, S. Fang, C. Ozsoy-Keskinbora, T. Suzuki, M. Kriener, L. Fu, E. Kaxiras, D. C. Bell, J. G. Checkelsky, Data for "Clean 2D superconductivity in a bulk van der Waals superlattice". Harvard Dataverse (2020); doi:10.7910/DVN/GSYYOA.

ACKNOWLEDGMENTS

We thank Y. Tokura, P. A. Lee, M. Nakano, H. Matsuoka, N. F. Q. Yuan, V. Mitrovic, and S.-L. Zheng for fruitful discussions, and M. Kamitani and A. Akey for technical support. **Funding:** Supported by the Gordon and Betty Moore Foundation through grants GBMF3848 and GBMF9070 to J.G.C. (instrumentation development); the Office of Naval Research (ONR) under award N00014-17-1-2883 (advanced characterization); the U.S. Department of Energy (DOE) Office of Science, Basic Energy Sciences, under award DE-SC0019300 (material development); the STC Center for Integrated Quantum Materials, NSF grant DMR-1231319 (A.D., S.F., C.O.-K., and E.K.); and the DOE Office of Basic Energy Sciences under award DE-SC0018945 (L.F.). Computations were performed on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant ACI-1548562. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by NSF Cooperative Agreement DMR-1157490, the State of Florida, and DOE. **Author contributions:** A.D. synthesized and characterized the single crystals. A.D. and H.I. performed the electrical transport experiments; A.D. and M.K. performed the magnetization experiments; C.O.-K. and D.C.B. performed the electronic microscopy experiments; A.D. and S.F. performed theoretical calculations; all authors contributed to discussions and writing of the manuscript; and J.G.C. coordinated the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data in the manuscript are available from the Harvard Dataverse (*45*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/231/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S26
Tables S1 to S6
References (*46*–*154*)

28 September 2019; accepted 21 August 2020
10.1126/science.aaz6643

VACCINES

Influenza vaccine–induced human bone marrow plasma cells decline within a year after vaccination

Carl W. Davis[1,2], Katherine J. L. Jackson[3]*, Megan M. McCausland[1,2], Jaime Darce[4]†, Cathy Chang[1,2], Susanne L. Linderman[1,2], Chakravarthy Chennareddy[1,2], Rebecca Gerkin[2,5], Shantoria J. Brown[2,5], Jens Wrammert[1,6], Aneesh K. Mehta[2,7], Wan Cheung Cheung[4]‡, Scott D. Boyd[3], Edmund K. Waller[2,5], Rafi Ahmed[1,2]§

A universal vaccine against influenza would ideally generate protective immune responses that are not only broadly reactive against multiple influenza strains but also long-lasting. Because long-term serum antibody levels are maintained by bone marrow plasma cells (BMPCs), we investigated the production and maintenance of these cells after influenza vaccination. We found increased numbers of influenza-specific BMPCs 4 weeks after immunization with the seasonal inactivated influenza vaccine, but numbers returned to near their prevaccination levels after 1 year. This decline was driven by the loss of BMPCs induced by the vaccine, whereas preexisting BMPCs were maintained. Our results suggest that most BMPCs generated by influenza vaccination in adults are short-lived. Designing strategies to enhance their persistence will be a key challenge for the next generation of influenza vaccines.

The goal of vaccination is to generate long-lasting protection against infection. Most vaccines in clinical use achieve this protection at least in part through the generation of pathogen-specific antibody responses. Antibody levels peak in the months after vaccination, followed by a decline to a plateau level that may be maintained for decades with minimal decline (*1*–*3*). Animal models have shown that these plateau antibody levels are maintained by nondividing, bone marrow–resident long-lived plasma cells (*4*, *5*). Studies of antibody synthesis rates have suggested that bone marrow plasma cells (BMPCs) are likely to produce most serum immunoglobulin G (IgG) in humans as well (*6*, *7*). Consistent with this, total and antigen-specific serum antibody levels correlate closely with BMPC numbers in humans (*8*, *9*). Antibody titers and protective immunity decline rapidly after seasonal influenza vaccination (*10*, *11*), suggesting either that the vaccine may fail to elicit BMPCs or that these BMPCs fail to become long-lived. To investigate these possibilities, we designed a clinical study of healthy adults (20 to 45 years of age) from whom we were able to obtain paired bone marrow samples before and 1 month after influenza vaccination, in conjunction with a full schedule of blood samples. In addition, we obtained a third bone marrow draw at ~1 year after vaccination from some of the subjects, allowing us to assess the long-term maintenance of vaccine responses in the bone marrow.

BMPC responses were measured in 53 volunteers who received the inactivated influenza vaccine between 2009 and 2018. Some of these 53 donors enrolled in multiple study years (see the materials and methods) and were treated as separate enrollments, for a total of 75 vaccine responses studied. In addition to providing blood samples on the day of vaccination (day 0) and on days 7, 14, 28, and 90 after vaccination, the volunteers also underwent bone marrow aspiration procedures on days 0 and 28. In addition, some donors provided blood and bone marrow samples ~1 year after vaccination, either at their reenrollment for the following year's study or as their final study time point.

We first examined influenza-specific immunity at baseline before vaccination. Using plasma cells enriched from the bone marrow aspirates, we performed enzyme-linked immunosorbent spot (ELISpot) assays to quantify influenza-specific BMPCs (Fig. 1A). Cells secreting influenza-specific antibodies (both IgG and IgA) were readily detectable in study volunteers before vaccination and were markedly more common than cells producing antibodies specific for tetanus toxin. Influenza-specific BMPC numbers were also five- to 10-fold higher than we have observed for varicella zoster virus (*12*). The percentage of influenza-specific IgG BMPCs on day 0 followed an approximately log-normal distribution (Fig. 1B). The geometric mean percentage of IgG-secreting BMPCs producing influenza-specific antibodies was 0.8%, with an interquartile range (IQR) of 0.52 to 1.3%. Across the cohort, the percentage of influenza-specific IgG^+ BMPCs correlated significantly with influenza-specific IgG titers in the blood (Fig. 1C). These data are consistent with a key role for BMPCs in maintaining humoral immunity to influenza. We also measured influenza-specific memory B cells (MBCs) in the blood of the donors. Influenza-specific cells accounted for 0.93% (IQR: 0.58 to 1.5%) of IgG MBCs in the blood (Fig. 1D). This MBC frequency correlated with blood antibody titers (Fig. 1E). Furthermore, there was a clear trend toward higher numbers of influenza-specific MBCs in those donors with the most influenza-specific BMPCs (Fig. 1F).

We next investigated whether seasonal influenza vaccination resulted in the production of new influenza-specific BMPCs. Influenza-specific BMPCs increased significantly to 1.9% (IQR: 1.2 to 2.9%) of IgG BMPCs by day 28 after vaccination (Fig. 2A and fig. S1A). The seasonal influenza vaccine contains hemagglutinin (HA) proteins from multiple viral strains: one from an H1N1 influenza A strain, one from an H3N2 strain, and either one or two HA proteins from influenza B strains. An HA protein similar or identical to that of the pandemic H1N1 strain A/California/7/2009 was included in the vaccine from the 2010–2011 season until the 2016–2017 season, allowing us to monitor responses against a single, consistent antigen in multiple seasons. IgG BMPCs specific for this HA increased from 0.11% (IQR: 0.07 to 0.18%) to 0.36% (IQR: 0.18 to 0.76%) on day 28 (Fig. 2A and fig. S1B). We also measured influenza-specific IgA BMPCs when plasma cell yields were high enough to run multiple antigen-specific ELISpot assays from each sample. Although we only obtained paired IgA data for six donors, there was a trend toward increasing numbers of IgA BMPCs (Fig. 2A). This trend was significant in the unpaired analysis (fig. S1C), which included additional donors for whom IgA BMPCs could be measured at day 0 or day 28 but not both. These increases were specific to plasma cells recognizing influenza vaccine antigens; no changes were observed in the percentage of cells secreting IgG specific for tetanus toxin (Fig. 2A and fig. S1C). Thus, we found no evidence of bystander effects of the vaccine on the frequencies of BMPCs specific for unrelated antigens.

We have previously shown that influenza-specific antibody-secreting cells (ASCs) appear in the blood at ~1 week after vaccination (*13*). Because bone marrow aspirate samples will unavoidably contain some level of contamination with peripheral blood, we wanted to confirm that influenza-specific ASCs were not in circulation at the time of bone marrow

[1]Emory Vaccine Center and Department of Microbiology and Immunology, Emory University, Atlanta, GA, USA. [2]Emory-UGA Center of Excellence of Influenza Research and Surveillance (CEIRS), Atlanta GA, USA. [3]Department of Pathology, Stanford University, Stanford, CA, USA. [4]Cell Signaling Technology, Inc., Danvers, MA, USA. [5]Department of Hematology and Oncology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA. [6]Department of Pediatrics, Division of Infectious Disease, Emory University School of Medicine, Atlanta, GA, USA. [7]Division of Infectious Diseases, School of Medicine, Emory University, Atlanta, GA, USA.
*Present address: Immunology Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia.
†Present address: Bluefin BioMedicine, Beverly, MA, USA.
‡Present address: BioMedicine Design, Medicinal Sciences, Worldwide Research & Development, Pfizer Inc., Cambridge, MA, USA.
§Corresponding author. Email: rahmed@emory.edu

aspiration. Influenza-specific IgG ASCs were at very low levels or undetectable in the blood on day 0, with the exception of one donor with a recent history of an influenza-like illness (green symbols). ASC numbers peaked at day 7, declined sharply by day 14, and were again very rare or undetectable on day 28 (Fig. 2B). Similar results were seen for IgA and IgM ASCs (fig. S1, D and E). Thus, the percentages of influenza-specific BMPCs on days 0 and 28 would not have been inflated by contamination with circulating ASCs. The magnitude of the day 7 ASC response correlated with the increase in influenza-specific BMPCs after vaccination (Fig. 2C). This is consistent with

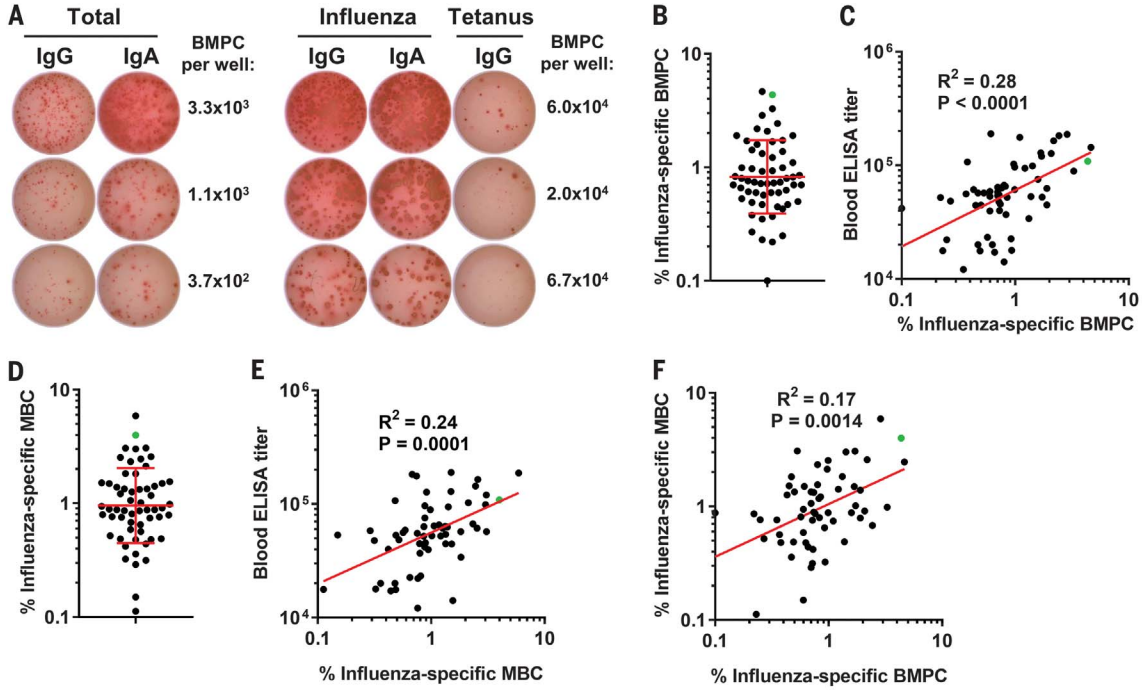


Fig. 1. Analysis of influenza-specific BMPCs before vaccination. (**A**) Example ELISpot data using magnetically enriched BMPCs with detection of cells secreting total or antigen-specific IgG or IgA. (**B**) Number of BMPCs producing influenza-specific IgG before vaccination expressed as a percentage of the total number of IgG-secreting BMPCs. (**C**) Correlation between prevaccination influenza-specific IgG BMPCs and influenza-specific IgG levels in the blood as measured by Enzyme-linked immunosorbent assay (ELISA). (**D**) Influenza-specific IgG MBCs from the peripheral blood are shown as a percentage of total IgG MBCs before vaccination. (**E**) Correlation between influenza-specific IgG MBCs and IgG titers in the blood before vaccination. (**F**) Correlation between the percentages of influenza-specific IgG BMPCs and MBCs. The data shown in (B) to (F) are for all study donors for which day 0 influenza-specific BMPC and MBC percentages were measured (n = 59 including repeat study participants; n = 43 unique donors). The means and standard deviations of the log-transformed data are indicated by red lines in (B) and (D). In (C), (E), and (F), Pearson correlation analysis was performed on the log-transformed BMPC, MBC, and ELISA data. The green dots represent a donor with evidence of recent influenza infection who was excluded from statistical calculations.

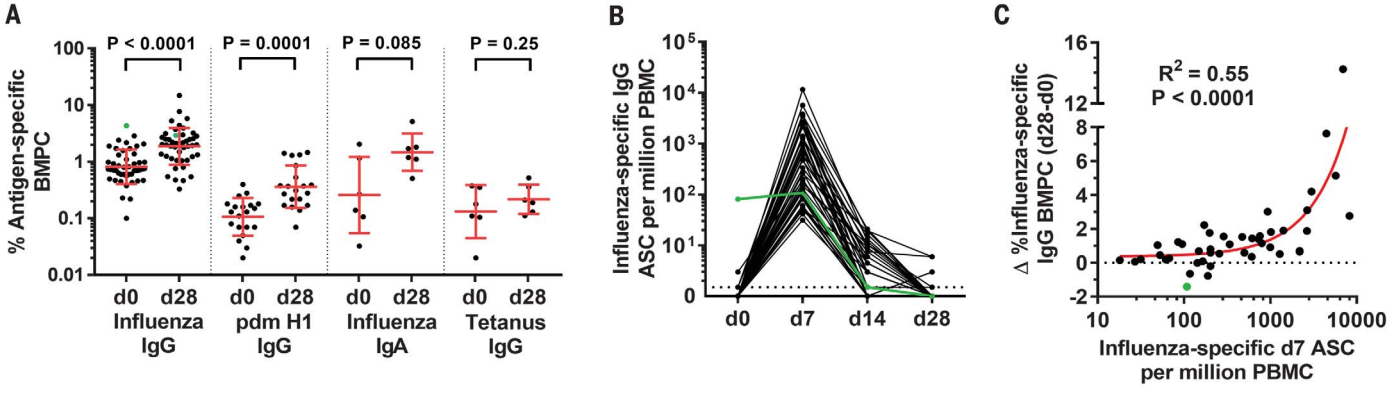


Fig. 2. Increase in influenza-specific plasma cells in the bone marrow after vaccination. (**A**) The percentage of IgG- or IgA-secreting BMPCs specific for the indicated antigens was measured by ELISpot on the day of vaccination (d0) and 28 days later (d28). Influenza, trivalent or quadrivalent inactivated influenza vaccine (n = 46 IgG responses measured from 34 unique donors; n = 6 IgA responses from six unique donors); pdm H1 HA, recombinant HA protein from pandemic H1N1 strain A/California/07/2009 (n = 20 responses from 14 unique donors); tetanus, tetanus toxin C fragment (n = 6 responses from six unique donors). P values shown are for paired, two-tailed t tests using the log-transformed data. The means and standard deviations of the log-transformed data are indicated in red. (**B**) Blood IgG ASC response to the influenza vaccine measured by ELISpot. Data are shown for n = 45 vaccine responses from 35 unique donors. The dotted line represents the lower limit of detection for the assay. (**C**) Pearson correlation between the magnitude of the IgG ASC response at day 7 and the increase in influenza-specific BMPCs, calculated as the percentage of influenza-specific IgG BMPCs at day 28 minus this percentage at day 0. Green symbols and lines in (A) to (C) indicate a donor with evidence of recent influenza infection who was excluded from the statistical analyses.

models suggesting that BMPCs are derived from blood ASCs that migrated to the bone marrow (*14*).

To determine whether the increase in influenza-specific BMPCs was maintained, we obtained bone marrow samples from a subset of donors at a late time point (mean 325 days after vaccination; range: 220 to 457 days). The percentage of influenza-specific BMPCs declined significantly between day 28 and the late sample (Fig. 3A). In subjects who reenrolled in multiple seasons and were vaccinated a second time, influenza-specific BMPCs typically increased after both immunizations, although there was substantial heterogeneity among donors (fig. S2). As shown in Fig. 3B, in the subset of donors who provided a long-term follow-up sample, there was a geometric mean 1.9-fold change (IQR: 1.3- to 3.2-fold) in influenza-specific BMPCs between days 0 and 28, similar to the overall cohort. However, by 1 year, the percentage of influenza-specific BMPCs had declined to baseline. Analysis of serum antibody levels revealed similar trends (Fig. 3C). The level of decline in BMPCs did not appear to correlate with the timing of the 1-year bone marrow aspiration (fig. S3), suggesting that the loss of newly generated influenza-specific BMPCs happened within 7 months after vaccination.

We next examined in detail the clonal composition of the B cell response to the vaccine by performing immunoglobulin heavy chain repertoire sequencing of BMPCs and ASCs from four donors. Three out of four donors (donors 35, 40, and 42) had a strong day 7 ASC response to the vaccine (fig. S4A), and all four donors showed serologic responses to each influenza strain in the vaccine (fig. S4B), particularly to the H1N1 component. To identify influenza-specific antibody families (or "clonotypes") in the repertoire sequencing data, we used two strategies: production of monoclonal antibodies (mAbs) cloned from day 7 ASCs (all four donors) and a proteomics-based approach in which antibodies specific to the H1 HA protein were identified from the polyclonal serum (donors 35 and 42) (fig. S5A). We identified 15 clonotypes that were highly expanded (>0.5% of day 7 ASC sequences) and determined to be influenza specific by either mAb cloning or proteomics analysis (fig. S5B and table S5). These antibody families already exhibited extensive somatic hypermutation in their heavy chain variable region genes at day 7, indicating that they were likely derived from preexisting MBCs. All 15 clonotypes were represented among both IgA- and IgG-expressing ASCs, although some clonotypes were enriched for one isotype or the other. We hypothesize that multiple MBCs from each clonotype responded to the vaccine, with varying recruitment of IgA versus IgG memory cells. Of the 11 H1-specific clonotypes that we identified, only one of them (clonotype 8 from donor 42) contributed detectably to H1-binding antibodies found in the serum on day 0 (fig. S5B). The other 10 H1-specific clonotypes could only be detected in the serum starting at day 7 or day 14. This suggests either that there were few long-lived plasma cells belonging to these clonotypes before vaccination or that these preexisting plasma cells produced antibodies that bound poorly to the pandemic H1 HA protein.

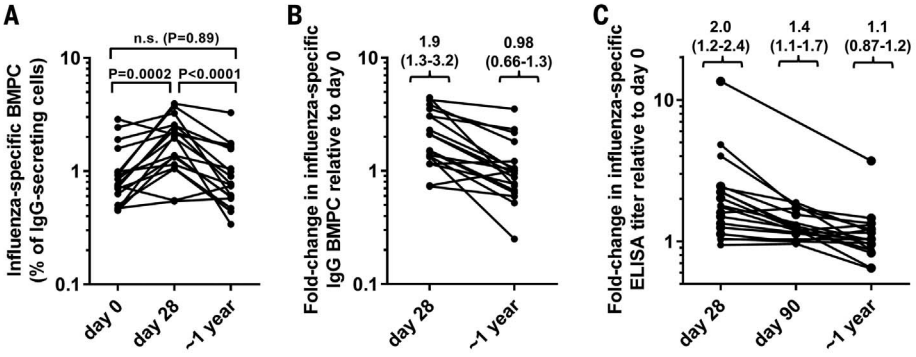


Fig. 3. Decline of influenza-specific BMPCs at 1 year after vaccination. (**A**) Percentage of influenza-specific IgG BMPCs shown over time in volunteers who returned for a third bone marrow aspiration ~1 year after vaccination. *P* values shown are for paired, two-tailed *t* tests of the log-transformed data. (**B**) Same data plotted as in (A) except the percentages of influenza-specific BMPCs at day 28 and 1 year are normalized to the day 0 percentage to visualize the fold change relative to prevaccination levels in each donor. Numbers above the points indicate the geometric means of the fold changes, with the IQRs in parentheses. (**C**) Influenza-specific blood IgG ELISA titers at day 28, day 90, and 1 year normalized to the day 0 titer and plotted as in (B). The same $n = 18$ responses from 15 unique donors are shown in all panels.

Of the 15 influenza-specific clonotypes that were expanded in the day 7 ASCs, 14 increased significantly among BMPC IgG sequences between days 0 and 28 after vaccination ($P = 0.0003$ by Wilcoxon signed-rank test; Fig. 4A). Thus, the same antibody families that contributed to the blood ASC response also contributed to the BMPC compartment, again consistent with the idea that blood ASCs give rise to BMPCs. Between day 28 and the late time point (~1 year), the percentage of BMPC IgG sequences belonging to these clonotypes decreased sharply ($P = 0.0017$). After 1 year, 11 of the 14 clonotypes still made up a higher percentage of BMPC IgG reads compared with day 0; however, this difference was not statistically significant ($P = 0.068$). There was a wide variation in how well the initial increase in frequency of each clonotype among BMPC sequences was sustained (Fig. 4B). The median increase in frequency for these 14 clonotypes after 1 year was only 0.15-fold as high as the initial increase seen at day 28 (IQR: 0.01- to 0.29-fold). In other words, for most clones, 70 to 99% of the newly generated BMPCs were lost over 1 year. As a separate strategy to measure clonotype abundances over time, we designed clone-specific quantitative polymerase chain reaction (qPCR) probe and primer sets (fig. S5C). Using this qPCR assay, we observed a similar trend where vaccine-specific antibody clonotypes expanded and then declined in BMPC RNA samples (median increase in frequency at 1 year was 0.09-fold of the initial increase at day 28; IQR: 0- to 0.38-fold). Both the initial increase in clonotype frequency among BMPCs and the long-term increase correlated with the representation of each clonotype among day 7 ASC sequences (Fig. 4, C and D), which is consistent with what we observed by ELISpot (Fig. 2C) and in keeping with the notion that blood ASCs give rise to BMPCs.

Our clonotype-tracking results suggested that the contraction in influenza-specific BMPCs between day 28 and 1 year that we saw by ELISpot was due to the loss of newly generated cells. To determine whether preexisting influenza-specific BMPCs were more stable over time, we examined "nonrecruited" influenza-specific clonotypes in the two donors whose H1-specific responses were profiled by serum proteomics analysis. These clonotypes were defined as HA-specific antibody families that were detected in the serum and in IgG BMPC sequences on day 0, but which represented <0.1% of day 7 vaccination-induced ASCs by antibody repertoire sequencing (tables S4 and S5). Most of these clonotypes could be detected among both IgG and IgA BMPC sequences (see table S5). The nonrecruited clones showed no significant change in their abundances over 1 year in either donor (Fig. 4, E and F), indicating that they likely represented a stable population of long-lived plasma cells that had been generated by previous influenza infection or vaccination.

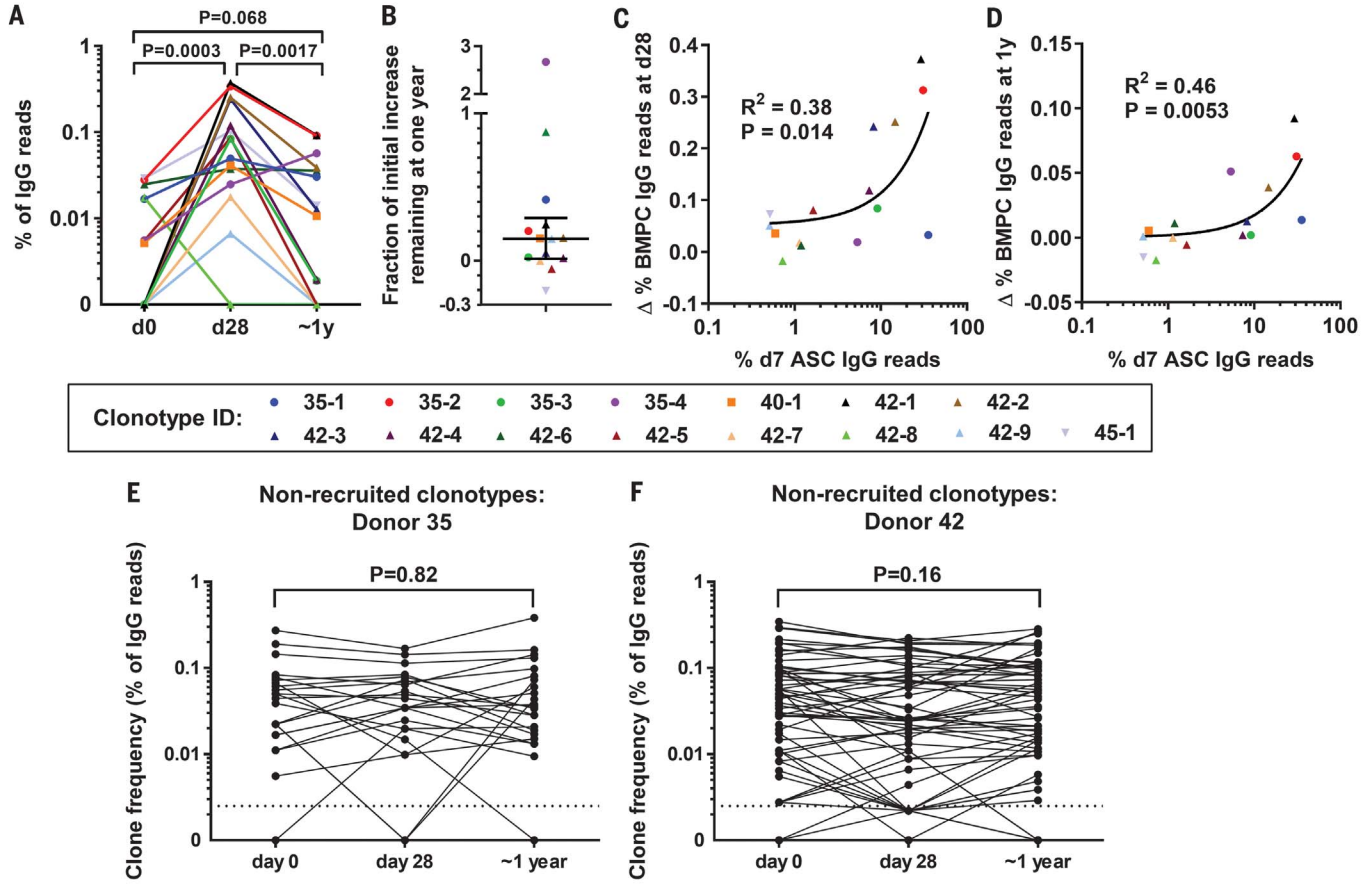


Fig. 4. Longitudinal tracking of influenza–specific antibody lineages (clonotypes) after vaccination. Clonotypes were tracked over time in BMPC and ASC immunoglobulin sequences. Fifteen clonotypes comprising at least 0.5% of day 7 blood ASC IgG sequences were considered to have been "recruited" into the vaccine-specific response. Clonotypes are named by their frequency among ASC sequences (e.g., "35-2" is the clonotype with second-highest read count in IgG sequences from day 7 ASCs from donor 35). (**A**) Percentage of BMPC IgG sequences belonging to each recruited clonotype shown on day 0, day 28, and 1 year after vaccination. (**B**) For the 14 recruited clonotypes that increased in the bone marrow between days 0 and 28, the long-term increase after 1 year is shown as a proportion of the initial increase. A clonotype that neither expanded nor contracted between day 28 and 1 year would have a value of 1 on this scale, whereas a clonotype that declined to its starting frequency would have a value of 0. Values >1 indicate that the clonotype expanded among BMPC sequences between day 28 and 1 year, and values below 0 indicate clonotypes that comprised a lower percentage of IgG reads at 1 year compared with day 0. (**C**) Correlation between the percentage of day 7 ASC IgG sequences belonging to each clonotype and the initial increase in that clonotype among BMPC IgG transcripts between days 0 and 28. (**D**) Correlation between the percentage of day 7 ASC IgG sequences belonging to each clonotype and the long-term increase of these clonotypes among BMPC sequences between day 0 and 1 year. (**E**) Persistence of preexisting, nonrecruited H1 HA-specific clonotypes in donor 35. Clonotypes were defined as preexisting if they were found in the BMPC repertoire sequencing data at day 0 and confirmed to be influenza specific by serum proteomics analysis or mAb production. Clonotypes were defined as nonrecruited if they accounted for <0.1% of day 7 IgG, IgM, and IgA ASC sequences. (**F**) Persistence of preexisting, nonrecruited H1 HA-specific clonotypes in donor 42. *P* values shown in (A), (E), and (F) are for the Wilcoxon signed-rank test. In (C) and (D), Pearson correlation was performed using the nontransformed data, and the best-fit linear regression lines are shown.

In recent years, it has become clear that humans are capable of generating protective antibody responses directed against epitopes on the influenza HA and neuraminidase (NA) proteins that are broadly conserved across diverse viral strains. This has raised hopes that a "universal" influenza vaccine targeting these conserved regions could be developed. If such a vaccine elicited protective antibody responses that persisted for at least several years, it would eliminate the need for annual immunization, which would be a major improvement on current vaccines. The data presented here add to our understanding of why humoral responses to influenza vaccine are not well maintained. We show that influenza-specific titers in the serum correlate with BMPC numbers before vaccination, supporting a key role for these cells in maintaining antibody levels. Vaccination does lead to the generation of new BMPCs, and some of these persist for at least 1 year. However, most of the newly generated BMPCs are lost within 1 year, showing that localization to the bone marrow is not enough to determine longevity. Two steps are required for a BMPC to become a long-lived plasma cell: (i) The cell must reach the appropriate survival niche and successfully compete for space there and (ii) it must undergo changes in gene expression and metabolism that promote longevity (*15*–*20*). Thus, our data suggest that most vaccine-induced BMPCs fail at one or both of these steps.

The decline in BMPC and antibody levels observed here was likely not specific to the influenza vaccine. One study of children receiving the hepatitis B vaccine showed declines in antibody titers of 85 to 90% from peak levels in the first year after vaccination (*21*). Conversely, in a study of adults given a first booster of the tetanus, diphtheria, and acellular pertussis vaccine (Tdap), IgG antibody levels to the

vaccine antigens declined only 40 to 66% in the first year and another 15 to 25% in the following year (*22*). This is a considerably slower rate of decay than we observed, suggesting that there is room for improvement with influenza vaccines.

A few findings from this study may be helpful for designing next-generation influenza vaccines capable of eliciting longer-term immunity. First, the blood ASC response after vaccination was highly predictive of the overall BMPC response, both in the short term and in the long term, and thus may be a useful marker for screening vaccine candidates. Because ASC generation after influenza vaccination can be increased through the use of adjuvants (*23*–*25*), it may be useful to assess whether adjuvants are able to improve long-term BMPC responses in humans. Second, most of the preexisting H1-specific antibody families detected in the serum on day 0 were not detected in the day 7 ASC response, and most of the H1-specific clonotypes that were recruited did not contribute appreciably to serum H1-specific titers before vaccination. This may reflect antibody feedback, in which preexisting serum antibodies binding to the vaccine block B cells recognizing the same epitopes from acquiring antigen (*26*–*30*). Designing strategies to overcome this effect will be critical for boosting long-term numbers of BMPCs targeting conserved epitopes.

REFERENCES AND NOTES

1. E. Hammarlund *et al*., *Nat. Med.* **9**, 1131–1137 (2003).
2. S. Crotty *et al*., *J. Immunol.* **171**, 4969–4973 (2003).
3. I. J. Amanna, N. E. Carlson, M. K. Slifka, *N. Engl. J. Med.* **357**, 1903–1915 (2007).
4. R. A. Manz, A. Thiel, A. Radbruch, *Nature* **388**, 133–134 (1997).
5. M. K. Slifka, R. Antia, J. K. Whitmire, R. Ahmed, *Immunity* **8**, 363–372 (1998).
6. R. McMillan *et al*., *J. Immunol.* **109**, 1386–1394 (1972).
7. T. Hibi, H. M. Dosch, *Eur. J. Immunol.* **16**, 139–145 (1986).
8. I. Turesson, *Acta Med. Scand.* **199**, 293–304 (1976).
9. T. Pritz *et al*., *Eur. J. Immunol.* **45**, 738–746 (2015).
10. G. T. Ray *et al*., *Clin. Infect. Dis.* **68**, 1623–1630 (2019).
11. P. F. Wright *et al*., *Pediatr. Infect. Dis. J.* **27**, 1004–1008 (2008).
12. C. S. Eberhardt *et al*., *J. Virol.* **94**, e02127-19 (2020).
13. J. Wrammert *et al*., *Nature* **453**, 667–671 (2008).
14. M. Odendahl *et al*., *Blood* **105**, 1614–1621 (2005).
15. N. Pengo *et al*., *Nat. Immunol.* **14**, 298–305 (2013).
16. L. D'Souza, D. Bhattacharya, *Immunol. Rev.* **288**, 161–177 (2019).
17. S. M. Lightman, A. Utley, K. P. Lee, *Front. Immunol.* **10**, 965 (2019).
18. J. L. Halliley *et al*., *Immunity* **43**, 132–145 (2015).
19. H. E. Mei *et al*., *Blood* **125**, 1739–1748 (2015).
20. A. Radbruch *et al*., *Nat. Rev. Immunol.* **6**, 741–750 (2006).
21. V. Gilca *et al*., *Vaccine* **27**, 6048–6053 (2009).
22. S. van der Lee *et al*., *Front. Immunol.* **9**, 681 (2018).
23. R. J. Cox *et al*., *PLOS ONE* **10**, e0131652 (2015).
24. L. Li, Y. Honda-Okubo, C. Li, D. Sajkov, N. Petrovsky, *PLOS ONE* **10**, e0132003 (2015).
25. A. H. Ellebedy *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17957–17964 (2020).
26. V. I. Zarnitsyna, J. Lavine, A. Ellebedy, R. Ahmed, R. Antia, *PLOS Pathog.* **12**, e1005692 (2016).
27. S. F. Andrews *et al*., *Immunity* **51**, 398–410.e5 (2019).
28. D. Hobson, F. A. Baker, R. L. Curry, *Lancet* **2**, 155–156 (1973).
29. Y. Zhang *et al*., *J. Exp. Med.* **210**, 457–464 (2013).
30. T. Arulraj, S. C. Binder, P. A. Robert, M. Meyer-Hermann, *Front. Immunol.* **10**, 2116 (2019).

ACKNOWLEDGMENTS

We thank the Emory Flow Cytometry Core for assistance with cell sorting, the Emory Integrated Genomics Core for assistance with MiSeq experiments, and R. Antia and V. Zarnitsyna for valuable discussions. **Funding:** NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS) contract HHSN272201400004C; Influenza Pathogenesis and Immunology Research Center (IPIRC, CEIRS) contract HHSN266200700006C; NIH grants 1R01AI127877 and 1R01AI130398 (SDB). **Author contributions:** Conceptualization: C.W.D., J.W., E.K.W., R.A.; Data curation: C.W.D., K.J.L.J., J.D., R.G., W.C.C.; Formal analysis: C.W.D., K.J.L.J.; Investigation: C.W.D., K.J.L.J., M.M.M., J.D., C.Che, C.Cha, S.L.L., W.C.C.; Methodology: C.W.D., K.J.L.J., M.M.M., J.D., J.W., W.C.C., S.D.B., E.K.W., R.A.; Project administration: E.K.W., R.A.; Funding acquisition: E.K.W., R.A.; Resources: K.J.L.J., J.D., R.G., W.C.C., A.K.M., S.D.B., E.K.W.; Software: K.J.L.J., S.D.B.; Supervision: W.C.C., S.D.B., E.K.W., R.A.; Visualization: C.W.D., K.J.L.J.; Writing – original draft: C.W.D., R.A.; Writing – Review & Editing: C.W.D., K.J.L.J., S.D.B., E.K.W., R.A. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Antibody high-throughput DNA-sequencing data have been deposited in the Sequence Read Archive (BioProject accession no. PRJNA649987).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/237/suppl/DC1
Materials and Methods
Tables S1 to S5
Figs. S1 to S5
References (*31*–*42*)

View/request a protocol for this paper from *Bio-protocol*.

16 October 2019; accepted 2 August 2020
Published online 13 August 2020
10.1126/science.aaz8432

CORONAVIRUS

MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses

Anna Bruchez[1]*†, Ky Sha[1]*, Joshua Johnson[2]‡, Li Chen[3], Caroline Stefani[1], Hannah McConnell[1]§, Lea Gaucherand[1]¶, Rachel Prins[1], Kenneth A. Matreyek[4]†, Adam J. Hume[5,6], Elke Mühlberger[5,6], Emmett V. Schmidt[7], Gene G. Olinger[2,5,6,8], Lynda M. Stuart[1,9], Adam Lacy-Hulbert[1,10]#

Recent outbreaks of Ebola virus (EBOV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have exposed our limited therapeutic options for such diseases and our poor understanding of the cellular mechanisms that block viral infections. Using a transposon-mediated gene-activation screen in human cells, we identify that the major histocompatibility complex (MHC) class II transactivator (CIITA) has antiviral activity against EBOV. CIITA induces resistance by activating expression of the p41 isoform of invariant chain CD74, which inhibits viral entry by blocking cathepsin-mediated processing of the Ebola glycoprotein. We further show that CD74 p41 can block the endosomal entry pathway of coronaviruses, including SARS-CoV-2. These data therefore implicate CIITA and CD74 in host defense against a range of viruses, and they identify an additional function of these proteins beyond their canonical roles in antigen presentation.

Recent and ongoing outbreaks of Ebola virus (EBOV) in Africa (*1*) and the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic highlight the need to identify additional treatment strategies for viral infections, including approaches that might complement traditional antivirals. Of particular interest is the identification of host-directed therapies that target common vulnerabilities and may be efficacious against multiple viruses, including those that may emerge in the future.

We set out to identify host pathways of cellular resistance to pathogens with pandemic potential, using a transposon-mutagenesis–forward genetic approach. We used a modified PiggyBac (PB) transposon (Fig. 1A), which stimulates or disrupts the expression of neighboring genes, thereby allowing an interrogation of both gene activation and inactivation in a single screen (*2*). Transposon-mutagenized libraries were treated with Ebola glycoprotein (EboGP)–expressing recombinant vesicular stomatitis virus (referred to as EboGP-VSV). Susceptible wild-type U2OS cells died after 3 to 4 days of treatment, whereas surviving cells could be expanded from mutagenized libraries and exhibited stable resistance to rechallenge with EboGP-VSV (Fig. 1B). These cells showed no cross-resistance to vesicular stomatitis virus (VSV) containing the VSV glycoprotein (VSVg-VSV) (Fig. 1C), which suggests that most of the resistance mechanisms selected in this screen targeted EboGP-mediated entry.

We identified candidate resistance genes by identifying genomic regions with high numbers of transposon insertions [referred to as common insertion sites (CISs)] (*3*). Combining data from eight independent screens revealed seven genomic loci with highly statistically significant ($P < 10^{-8}$) CISs that occurred in more than one screen, representing high-confidence candidate-resistance mutations (Fig. 1D, outer ring). Likely target genes of transposon insertions were identified on the basis of transposon insertion position and orientation (Fig. 1D and table S1). We focused on the two genes that were found in all eight screens using the most stringent criteria.

The first of these was *NPC1*, located on chromosome 18. All transposon insertions at this site were intragenic in both sense and antisense orientations, and all were predicted to disrupt *NPC1* expression (Fig. 1E). This is consistent with the role of *NPC1* as the EBOV receptor (*4*, *5*) and validates our screening approach. Notably, U2OS cells are haploid at the *NPC1* locus (*6*), and these transposon insertions are therefore predicted to generate *NPC1*-null cells, which explains why *NPC1* was the only predicted gene-disruption mutant identified as a high-stringency candidate gene.

All transposon insertions at the second CIS—located on chromosome 16—were upstream of the gene *CIITA* and were oriented in the sense orientation, consistent with activation of expression (Fig. 1F and fig. S1). CIITA overexpression in wild-type U2OS cells increased cell survival, reduced green fluorescent protein (GFP) reporter expression, and completely inhibited plaque formation, which confirms that CIITA increases resistance to EboGP-VSV 100- to 1000-fold (Fig. 2, A to E, and fig. S2). CIITA-overexpressing cells were also resistant to EboGP-pseudotyped single cycle viruses (Fig. 2, F and G), which strongly suggests that CIITA inhibits viral entry rather than targeting viral transactivators as suggested for HIV and human T cell leukemia virus (HTLV) (*7*, *8*). Furthermore, using EboGP virus–like particles (EboGP-VLPs) carrying β-lactamase (*9*), we found that CIITA did not affect the internalization of EboGP-VLPs into cells (Fig. 2H), but it blocked viral fusion, which occurs in the endosome (*10*) (Fig. 2I). CIITA-expressing U2OS cells were also highly resistant to infection by high titers of native EBOV, showing reduced reporter gene expression, cell death, and plaque formation (Fig. 2, J to M). CIITA expression did not inhibit replication of an EBOV minigenome, which indicates that CIITA does not act on the viral replication complex (fig. S3). Furthermore, CIITA inhibited infection mediated by glycoproteins (GPs) from a range of EBOV species—including Sudan, Zaire, and Reston—as well as by those from the distantly related filovirus Marburg virus (Fig. 2G). Thus, CIITA induces broad antiviral activity against EBOV and other pathogenic filoviruses through the inhibition of viral GP-mediated entry.

CIITA, also known as NLRA, is a nucleotide-binding oligomerization domain (Nod)–like receptor (NLR) (*11*), but unlike most other NLRs, which function as cytosolic sensors, CIITA is a transcription factor (*12*). We therefore hypothesized that its antiviral activity occurred through the altered expression of host target genes. Supporting this hypothesis, mutation of domains required for transcriptional activity completely ablated CIITA antiviral activity (fig. S4). Resistance also required NF-Y, a component of the enhanceosome multiprotein complex, which mediates transcriptional activation by CIITA (*13*), but resistance was independent of another enhanceosome protein, RFX5 (figs. S4 and S5). Antiviral activity was therefore mediated by a subset of NF-Y–dependent, RFX5-independent CIITA target genes, which includes genes associated with antiviral immunity (*14*). Systematic knockdown of all CIITA target genes identified a single gene, *CD74*, required for CIITA-mediated resistance (Fig. 3, A and B). This was confirmed

[1]Benaroya Research Institute, Seattle, WA 98101, USA. [2]National Institute of Allergy and Infectious Diseases (NIAID) Integrated Research Facility, Frederick, MD 21702, USA. [3]Massachusetts General Hospital, Boston, MA 02114, USA. [4]Department of Genome Sciences, University of Washington, Seattle, WA 98109, USA. [5]Boston University School of Medicine, Boston, MA 02118, USA. [6]National Emerging Infectious Diseases Laboratories, Boston University, Boston, MA 02118, USA. [7]Merck and Co., Inc, Kenilworth, NJ 07033, USA. [8]MRIGlobal, Gaithersburg, MD 20878, USA. [9]Bill and Melinda Gates Foundation, Seattle, WA 98109, USA. [10]Department of Immunology, University of Washington, Seattle, WA 98109, USA.
*These authors contributed equally to this work.
†Present address: Department of Pathology, Case Western Reserve University, Cleveland, OH 44106, USA.
‡Present address: AbViro LLC, Bethesda, MD 20814, USA.
§Present address: Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.
¶Present address: Graduate Program in Molecular Microbiology, Tufts Graduate School of Biomedical Sciences and Department of Molecular Biology and Microbiology, Tufts University, Boston, MA 02155, USA.
#Corresponding author. Email: alacyhulbert@benaroyaresearch.org

by CRISPR knockout of CD74 expression and function in CIITA-overexpressing cells (Fig. 3C).

Both CIITA and CD74 are expressed at high levels by macrophages and dendritic cells (DCs), which are early targets of EBOV (*15*, *16*). To test whether CIITA has antiviral activity in immune cells, we used primary bone marrow–derived macrophages (BMDMs) from *Ciita*$^{-/-}$ and *Cd74*$^{-/-}$ mice. Naïve BMDMs did not express high levels of CIITA or CD74, and they showed no difference in viral fusion. Treatment with interferon-γ (IFN-γ) and lipopolysaccharide (LPS) induced expression of CIITA and CD74, and *Ciita*$^{-/-}$ and *Cd74*$^{-/-}$ BMDMs primed with IFN-γ and LPS had higher levels of EboGP-VLP fusion than those observed in equivalent wild-type cells (Fig. 3, D to G, and fig. S6). Similar results were seen in *Cd74*$^{-/-}$ bone marrow–derived DCs and in a *CD74*$^{-/-}$ human macrophage-like cell line (differentiated THP-1) (figs. S7 and S8). Thus, endogenous CIITA and CD74 have antiviral activity in primary immune cells, which can be induced by exposure to IFN-γ and LPS.

CD74 is the major histocompatibility complex class II (MHC-II) invariant chain, and human cells express four main isoforms of CD74, which differ in the presence of an N-terminal endoplasmic reticulum (ER) retention signal and an internal thyroglobulin domain (Fig. 4A) (*17*). Only one CD74 isoform, p41, was able to fully rescue resistance to EboGP-VSV infection in CIITA-expressing,

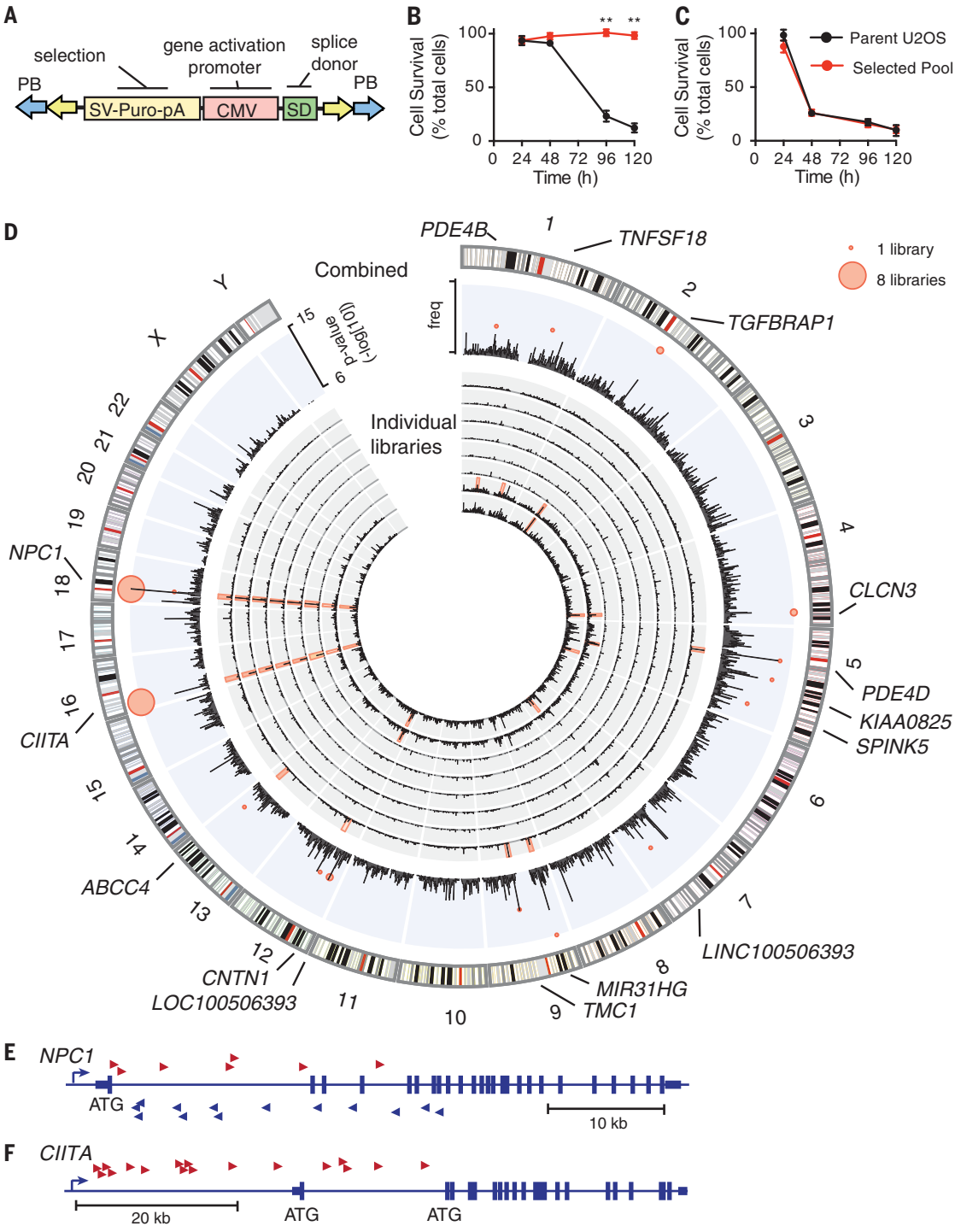


Fig. 1. Transposon-mediated activation tagging generates mutant cells resistant to Ebola. (**A**) Modified PB transposon. SV-Puro-pA, puromycin selection cassette; CMV, cytomegalovirus promoter; SD, splice donor. (**B** and **C**) Resistance of selected cells to EboGP-VSV (B) and VSVg-VSV (C). Data are means ± SD of *n* = 3 replicates for one representative pool. Student's *t* test; $^{**}P < 0.01$. (**D**) Distribution of transposon insertions. Inner rings show insertions per 1 Mb for individual libraries (black histograms) and CISs ($P < 10^{-7}$). Outer ring shows combined insertions for all libraries (black histogram) and lowest *P* value for CISs (red bubble plot). Point size represents the number of libraries with the CIS. freq, frequency. (**E** and **F**) Cumulative independent insertions from all eight libraries mapping to *NPC1* (E) and *CIITA* (F).

CD74-knockout cells (Fig. 4B and fig. S9). p41 conferred resistance independently of CIITA expression (Fig. 4C), which demonstrates that CD74 p41 expression was sufficient to induce antiviral activity. This property of CD74 was not limited to U2OS cells, as CD74 p41 similarly inhibited fusion when expressed in THP-1 cells (Fig. 4D). The p41 isoform contains the thyroglobulin domain, lacks the ER retention signal, and normally accumulates in endosomes. Mutant constructs of CD74 revealed that only the thyroglobulin domain is essential for antiviral activity, but dissociation from the membrane—either by addition of a furin cleavage site (labeled furin in Fig. 4E)

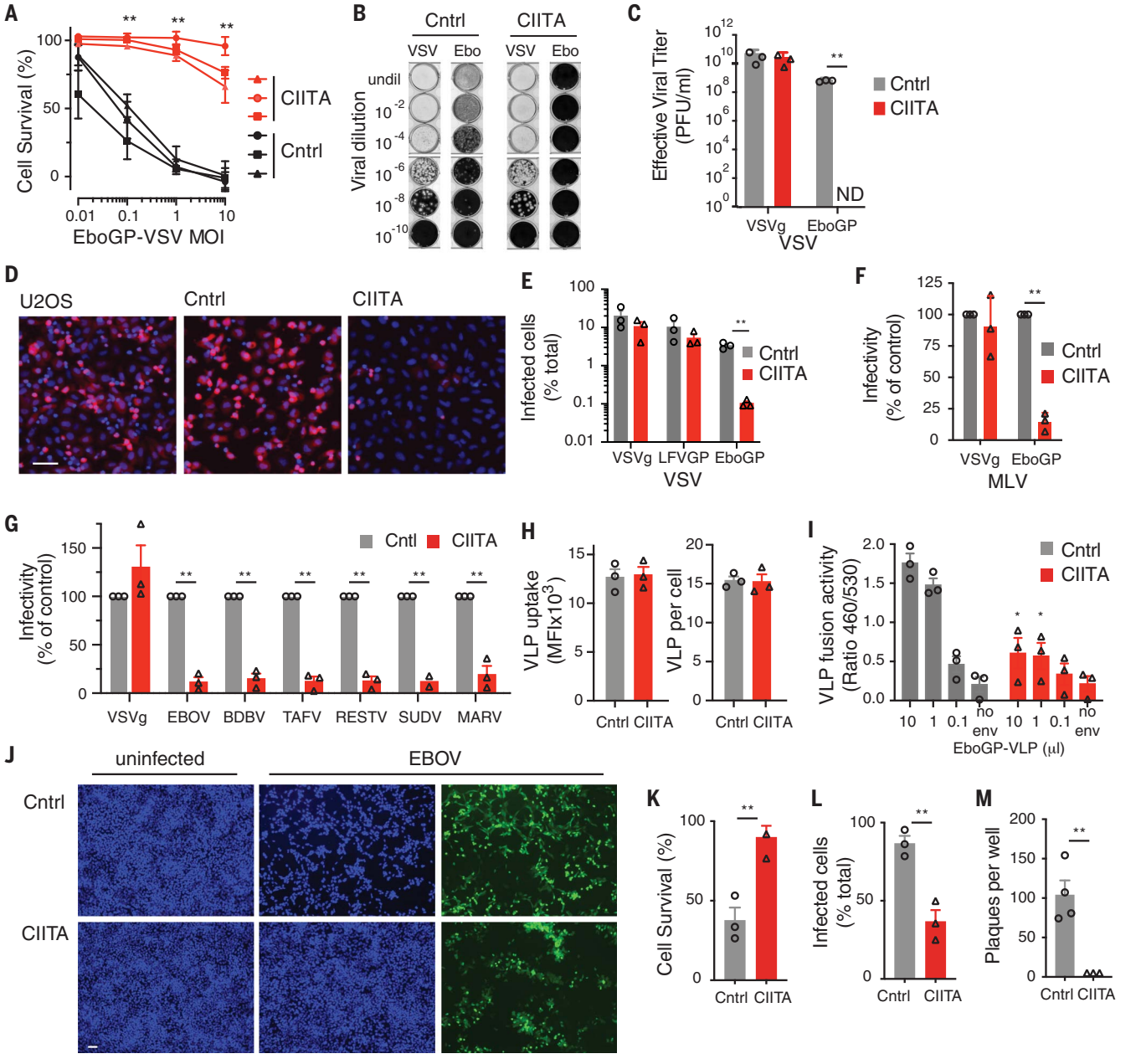


Fig. 2. Identification of CIITA as an Ebola restriction factor. (**A**) Resistance of CIITA-overexpressing and control (Cntrl) U2OS cells EboGP-VSV. MOI, multiplicity of infection. (**B** and **C**) Plaque formation assays (B) and effective viral titer (C) for control and CIITA-overexpressing U2OS cells infected with VSVg-VSV (VSV) and EboGP-VSV (Ebo). undil, undiluted; PFU, plaque-forming units. (**D**) Representative images of CIITA-transfected (CIITA), control-transfected (Cntrl), and unmanipulated U2OS cells (U2OS) infected with mCherry-expressing EboGP-VSV (red) and stained with Hoechst 33342 to resolve cell nuclei (blue). (**E** to **G**) Infection of control and CIITA-expressing U2OS cells by recombinant VSV pseudotyped with EboGP, LFVGP (Lassa virus GP), or VSVg (E); single cycle murine leukemia virus (MLV) pseudotyped with VSVg and EboGP (F); or single cycle HIV pseudotyped with VSVg or GP from EBOV, Taï Forest virus (TAFV), Bundibugyo virus (BDBV), Sudan virus (SUDV), Reston virus (RESTV), or Marburg virus (MARV) (G). (**H** and **I**) Internalization (H) and fusion (I) of EboGP-VLPs by control and CIITA-overexpressing U2OS cells. No env, nonenveloped control VLPs. (**J** to **M**) Infection of control and CIITA-overexpressing U2OS cells by infectious EBOV measured by imaging of GFP reporter (green) and cell nuclei (blue) (J), cell survival (K), infected cells (L), or plaque formation (M). Data are means ± SEM of three independent experiments [(A) to (I)] or experiments with three independent cell clones [(K) to (M)]. Student's *t* test [(A), (C), and (K) to (M)] or analysis of variance (ANOVA) with Tukey's multiple comparison test [(E) to (I)]; $*P < 0.05$; $**P < 0.01$; ND, not detected. Scale bars, 100 µm.

or deletion of the transmembrane sequence (No TM in Fig. 4E)—or delivery to the cell surface by fusion to a heterologous cytoplasmic and transmembrane sequence from tetherin (tetherin in Fig. 4E) almost completely removed antiviral activity (Fig. 4E and fig. S10). Thus, antiviral activity required delivery of the thyroglobulin domain to the endosomal membrane. Electron microscopy showed that EboGP-VSV virions accumulated in late endosomal multivesicular bodies (MVBs) of CIITA- and CD74 p41–expressing cells, with some virions within intraluminal vesicles (Fig. 4F and fig. S11). Confocal microscopy confirmed that virus-like particles (VLPs) localized proximal to CD63 and the ESCRT component Hrs, which mark MVBs (*18*, *19*) (Fig. 4, G and H). Thus, CIITA and CD74 p41 inhibit fusion by arresting viral particles in MVB compartments.

EBOV entry requires endosomal cathepsins (*4*, *10*, *20*) (fig. S12), which sequentially process EboGP (Fig. 4I and fig. S13). The CD74 thyroglobulin domain inhibits cathepsins (*21*), which suggests that this may be the mechanism for antiviral activity. In support of this, CD74 inhibited EboGP processing, similar to the effects of the cathepsin L (CTSL) inhibitor FYDMK (Fig. 4I). Additionally, disruption of the p41 CTSL binding site (*22*, *23*) by mutation completely inhibited antiviral activity (Fig. 4J and fig. S10). GP cleavage by endosomal proteases facilitates the entry of other viruses, including coronaviruses. SARS-CoV and SARS-CoV-2 S proteins can be processed by either endosomal cathepsin B and CTSL or alternatively by cell-surface serine proteases including TMPRSS2 (*24*, *25*). In TMPRSS-expressing cells, such as lung epithelium, inhibition of both cathepsins and serine proteases is required to inhibit viral entry, whereas cathepsin inhibitors alone block infection in cell lines—such as U2OS and Vero cells—that lack

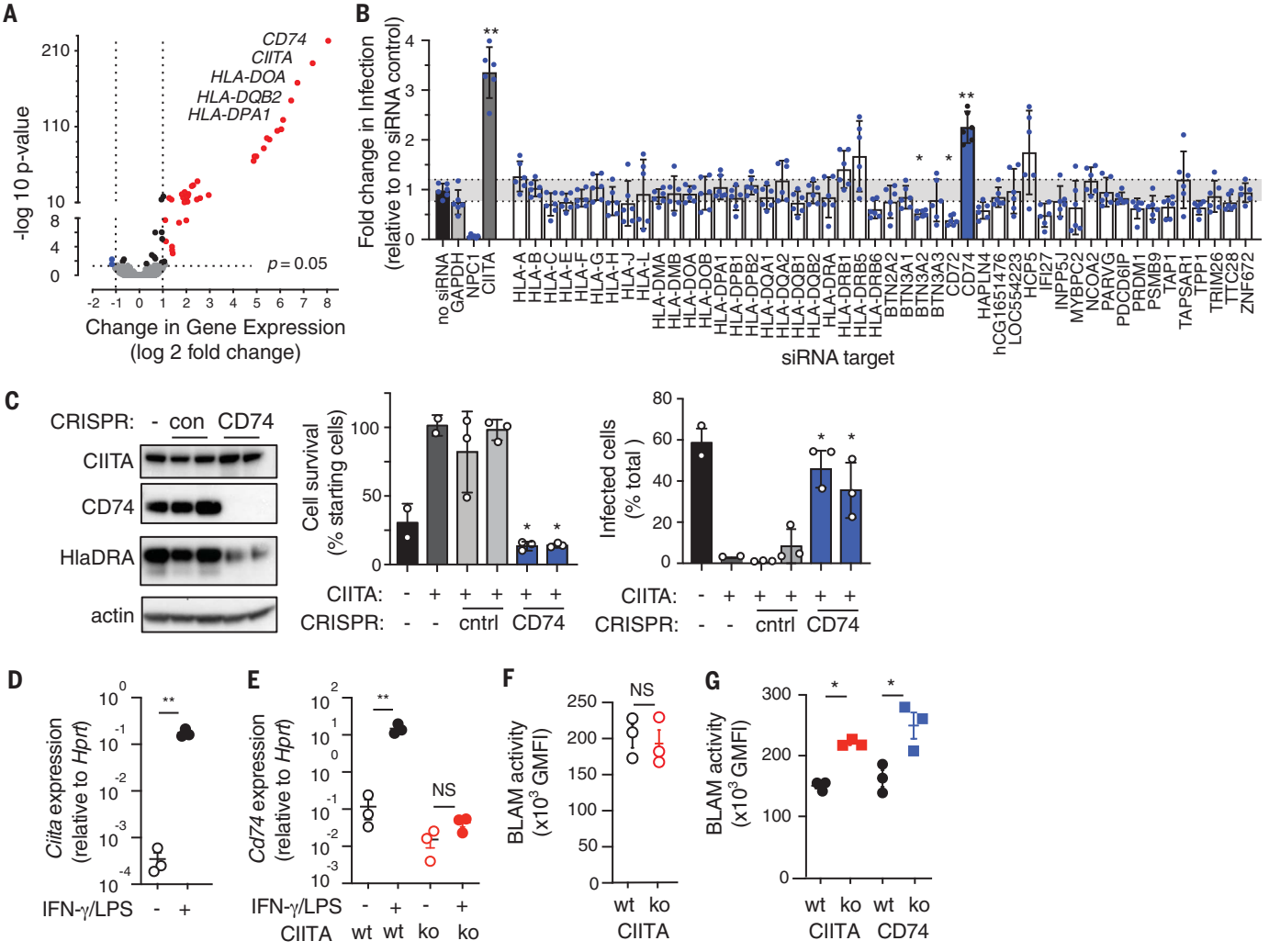


Fig. 3. Transcriptional activity of CIITA and enhanceosome components are required for resistance. (**A**) Genes regulated by CIITA in U2OS cells, with strongest induced genes identified. Mean of three independent CIITA-expressing clones and controls. (**B**) EboGP-VSV infection of CIITA-expressing cells treated with small interfering RNA (siRNA) against CIITA transcriptional targets. Data are from two siRNAs per gene, $N = 3$ independent screens, and bars indicate means with 95% confidence intervals (CIs). One-way ANOVA with Bonferroni's multiple comparisons; $^{*}P < 0.05$; $^{**}P < 0.01$. Dotted lines indicate 99% CIs from no siRNA control. (**C**) *CD74* CRISPR-targeting in CIITA-overexpressing U2OS cells was verified by immunoblot, and infection and survival were measured after EboGP-VSV challenge. Data are means ± SEM of $N = 3$ experiments using two independent cell clones. (**D** and **E**) *Ciita* and *Cd74* expression in wild-type (wt) or *Ciita*$^{-/-}$ mouse BMDMs with or without priming by IFN-γ and LPS. ko, knockout; NS, not significant. (**F** and **G**) Fusion of EboGP-VLPs in unprimed (F) or primed (G) mouse BMDM from *Ciita*$^{-/-}$ and *Cd74*$^{-/-}$ mice, measured as geometric mean fluorescence (GMFI) of cleaved CCF2. BLAM, β-lactamase. Data are means ± SEM for independent cultures from three mice per group. Student's *t* test; $^{*}P < 0.05$; $^{**}P < 0.01$. Similar results were observed in three independent experiments.

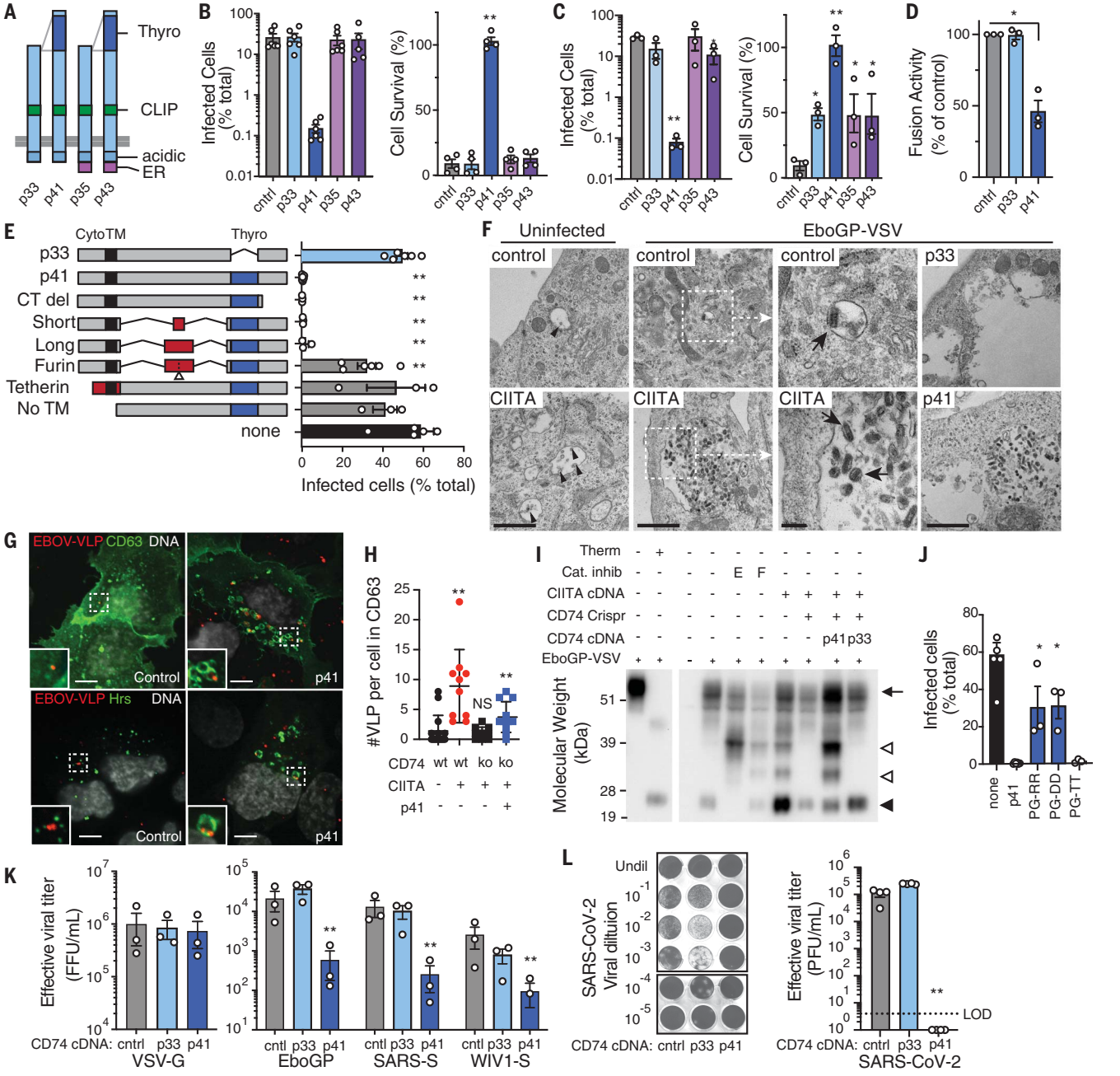


Fig. 4. CD74 p41 inhibits cathepsin-mediated cleavage of EboGP. (**A**) Human CD74 isoforms with ER retention signal (ER), CLIP, acidic, and p41 thyroglobulin (Thyro) domains. (**B** and **C**) EboGP-VSV infection and survival of *Cd74*$^{-/-}$ CIITA-expressing (B) or wt (C) U2OS cells expressing CD74 isoforms. (**D**) EboGP-VLP fusion in THP-1 macrophage-like cells expressing CD74 p33 and p41. (**E**) EboGP-VSV infection of U2OS cells expressing CD74 mutant constructs. Cyto, cytoplasmic domain; TM, transmembrane domain; Thyro, thyroglobulin domain; CT del, carboxy-terminus deletion; No TM, deletion of the transmembrane sequence. (**F**) Transmission electron micrographs of control, CIITA-expressing, and CD74-expressing U2OS cells 3 hours after infection with EboGP-VSV. Dotted-line regions are enlarged in adjacent panels (as indicated by white arrows). Intraluminal vesicles (black arrowheads) and internalized EboGP-VSV (black arrows) are marked. Scale bars, 1 μm (left, center left, and right panels) and 200 nm (center right panels). (**G**) Confocal microscopy of control and p41-expressing U2OS cells showing EBOV-VLP (red), CD63, or Hrs (green), and nuclei (white). Scale bars, 10 μm. (**H**) VLPs associated with CD63 endosomes in U2OS cells expressing CIITA and CD74 as indicated. Each point represents a single cell, mean ± SD $n \geq 9$. Mann-Whitney *U* test; $^{**}P < 0.01$. Similar results were seen in three independent experiments. (**I**) Immunoblot of EboGP in EboGP-VSV–infected U2OS cells. EboGP-VSV preparation ± thermolysin (Therm) is shown for reference (left). Cells were treated with cathepsin inhibitors (Cat. inhib) E64D (E) or FYDMK (F), or expressed CIITA and CD74. EboGP in virus particles (arrow), after proteolysis (closed arrowhead), and after partial cleavage (open arrowhead) are indicated. (**J**) EboGP-VSV infection of U2OS cells expressing p41 with CTSL binding site mutations. (**K**) Infection of control, p33-, or p41-expressing U2OS cells by HIV-GFP pseudotyped with GPs from VSV, EBOV, SARS-CoV, or WIV1-CoV, measured as focus-forming units per milliliter of virus (FFU/ml). (**L**) Infection of control, p33-, or p41-expressing Vero cells by SARS-CoV-2, showing representative crystal violet-stained monolayers and infection measured as plaque-forming units per milliliter of virus (PFU/ml). Except where indicated, data are means ± SEM of data from ≥3 independent experiments. Student's *t* test with Benjamini correction; $^{*}P < 0.05$; $^{**}P < 0.01$.

TMPRSS2 (*25*). p41 inhibited the entry of viruses pseudotyped with S proteins from SARS-CoV and a related bat virus, WIV1-CoV, into U2OS cells, which demonstrates that p41 inhibited S protein processing (Fig. 4K). To determine whether p41 exhibited antiviral activity against authentic SARS coronavirus, we challenged p41-expressing Vero E6 cells with SARS-CoV-2. CD74 p41 expression completely inhibited plaque formation, which demonstrates that this antiviral activity extended beyond filoviruses (Fig. 4L).

Here, we identify the antiviral activity of CIITA and CD74. We show that CIITA induces resistance by up-regulation of the p41 isoform of CD74, which blocks cathepsin-mediated cleavage of viral GPs, thereby preventing viral fusion. This antiviral activity protects against a wide range of cathepsin-dependent viruses, including filoviruses and coronaviruses; functions in macrophages and DCs that are early targets of infection (*15*, *16*); and is activated by IFN-γ. We demonstrate that CIITA and CD74 mediate the endosomal sequestration of certain viruses as a mechanism of cellular host defense. We speculate that this activity is evolutionarily ancient and precedes their better-known role in antigen processing. We anticipate that the application of this transposon screening approach to other models of infection will reveal additional mechanisms that have eluded conventional screening strategies.

REFERENCES AND NOTES

1. S. K. Gire *et al*., *Science* **345**, 1369–1372 (2014).
2. L. Chen *et al*., *BMC Cancer* **13**, 93 (2013).
3. T. L. Bergemann *et al*., *Nucleic Acids Res.* **40**, 3822–3833 (2012).
4. M. Côté *et al*., *Nature* **477**, 344–348 (2011).
5. J. E. Carette *et al*., *Nature* **477**, 340–343 (2011).
6. J. Barretina *et al*., *Nature* **483**, 603–607 (2012).
7. H. Okamoto, K. Asamitsu, H. Nishimura, N. Kamatani, T. Okamoto, *Biochem. Biophys. Res. Commun.* **279**, 494–499 (2000).
8. G. Tosi *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12861–12866 (2006).
9. C. J. Shoemaker *et al*., *PLOS ONE* **8**, e56265 (2013).
10. J. S. Spence, T. B. Krause, E. Mittler, R. K. Jangra, K. Chandran, *mBio* **7**, e01857-15 (2016).
11. J. A. Harton, M. W. Linhoff, J. Zhang, J. P. Y. Ting, *J. Immunol.* **169**, 4088–4093 (2002).
12. R. S. Accolla *et al*., *J. Exp. Med.* **164**, 369–374 (1986).
13. X. S. Zhu *et al*., *Mol. Cell. Biol.* **20**, 6051–6061 (2000).
14. D. Wong *et al*., *Genome Biol.* **15**, 494 (2014).
15. T. W. Geisbert *et al*., *Am. J. Pathol.* **163**, 2347–2370 (2003).
16. M. Bray, T. W. Geisbert, *Int. J. Biochem. Cell Biol.* **37**, 1560–1566 (2005).
17. M. Strubin, C. Berte, B. Mach, *EMBO J.* **5**, 3483–3488 (1986).
18. T. Kobayashi *et al*., *Mol. Biol. Cell* **11**, 1829–1843 (2000).
19. K. G. Bache, A. Brech, A. Mehlum, H. Stenmark, *J. Cell Biol.* **162**, 435–442 (2003).
20. E. H. Miller *et al*., *EMBO J.* **31**, 1947–1960 (2012).
21. M. Mihelič, A. Dobersek, G. Guncar, D. Turk, *J. Biol. Chem.* **283**, 14453–14460 (2008).
22. G. Gunčar, G. Pungerčič, I. Klemenčič, V. Turk, D. Turk, *EMBO J.* **18**, 793–803 (1999).
23. A. M. Lennon-Duménil *et al*., *EMBO J.* **20**, 4055–4064 (2001).
24. G. Simmons *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11876–11881 (2005).
25. M. Hoffmann *et al*., *Cell* **181**, 271–280.e8 (2020).

ACKNOWLEDGMENTS

We thank M. Mason, M. Rosasco, S. Presnell, and the Bioinformatics Department at Benaroya Research Institute (BRI) for support in data analysis and V. Gersuk and the BRI genomics core for sequencing. We thank B. Schneider and S. MacFarlane from the Electron Microscopy Resource at Fred Hutch for help with transmission electron microscopy experiments and L. Eisenlohr and M. O'Mara at Children's Hospital of Philadelphia for providing *Cd74*-knockout mouse bone marrow. **Funding:** This work was supported by National Institutes of Health grants R33AI102266, U01AI070330, and R33AI119341 (to A.L.-H. and L.M.S.); U19AI125378-04S1 (to A.L.-H.); and R21AI135912 (to E.M.). Work at NIAID Integrated Research Facility was funded by contract no. HHSN272200700016I to Battelle Memorial Institute (BMI). J.J. performed this work as an employee of BMI. SARS-CoV-2 work was performed in the BSL3 at Case Western Reserve University (CWRU), which is supported by the CWRU and University Hospitals Center for AIDS research grant P30AI36219. **Author contributions:** A.B. performed most of the experiments. Screen and data analysis tools were developed by K.S. BSL4 experiments were performed by J.J. and G.G.O., minigenome experiments were performed by A.J.H. and E.M., and K.A.M. designed all CD74 and CIITA mutations. H.M., R.P., and L.G. provided technical assistance. C.S. assisted with data analysis and visualization. G.G.O. and E.M. provided assistance with experimental planning and data interpretation. L.C., E.V.S., L.M.S., and A.L.-H. conceived the study. The manuscript was written by A.L.-H. and L.M.S. with assistance from A.B. and K.S. **Competing interests:** E.V.S. is presently an employee of Merck and Co., Inc., Kenilworth, NJ, and holds stock in Merck and Co. This work was conducted before E.V.S.'s affiliation with Merck. The authors declare no other competing interests. **Data and materials availability:** Full analysis of screen results is presented in the supplementary materials. DNA and RNA sequencing data are deposited at Gene Expression Omnibus (under accession nos. GSE156598 and GSE155204, respectively). The PB transposon was obtained under a material transfer agreement with the Wellcome Trust Sanger Institute. All other data are available in the manuscript or the supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/241/suppl/DC1
Materials and Methods
Figs. S1 to S14
Tables S1 to S5
References (*26*–*43*)

View/request a protocol for this paper from *Bio-protocol*.

21 February 2020; accepted 20 August 2020
Published online 27 August 2020
10.1126/science.abb3753

NEUROSCIENCE

Alternating sequences of future and past behavior encoded within hippocampal theta oscillations

Mengni Wang[1], David J. Foster[2], Brad E. Pfeiffer[1]*

Neural networks display the ability to transform forward-ordered activity patterns into reverse-ordered, retrospective sequences. The mechanisms underlying this transformation remain unknown. We discovered that, during active navigation, rat hippocampal CA1 place cell ensembles are inherently organized to produce independent forward- and reverse-ordered sequences within individual theta oscillations. This finding may provide a circuit-level basis for retrospective evaluation and storage during ongoing behavior. Theta phase procession arose in a minority of place cells, many of which displayed two preferred firing phases in theta oscillations and preferentially participated in reverse replay during subsequent rest. These findings reveal an unexpected aspect of theta-based hippocampal encoding and provide a biological mechanism to support the expression of reverse-ordered sequences.

Experience necessarily occurs in a sequential manner, and hippocampal function is critical for representation and storage of sequential information. Adaptive behavior requires the ability to analyze experience, both prospectively and retrospectively. It remains unclear how forward-ordered neural activity can facilitate storage or expression of reverse-ordered sequences, which are observed in ripple-based reverse replay (*1*, *2*) and may underlie human episodic memory retrieval (*3*). Accurate, precise, and stable spatial representation across the hippocampal network of place cells is insufficient to support goal-directed spatial navigation in the absence of theta oscillations (*4*, *5*). This indicates that the precise, population-level patterns of activity organized by theta oscillations, termed theta sequences (*6*–*9*), are critical for hippocampal-dependent, memory-guided behavior (*10*). We examined ensemble place cell activity recorded from bilateral hippocampal area CA1 in rats engaged in reward-seeking exploration of both linear tracks and open arenas (*11*, *12*). Large numbers of simultaneously monitored place cells (per-session mean ± SEM = 144.3 ± 14.0 putative excitatory units) allowed a memoryless, uniform-prior Bayesian decoding algorithm to accurately extract the spatial information encoded by the recorded ensemble in each session (table S1). We investigated the temporally compressed spatial trajectories expressed within theta oscillations during active exploration (speed ≥10 cm/s), examining a total of 115,845 theta oscillations (7240.3 ± 1126.7 per session).

We quantified sequential spatial representation within theta sequences by aligning each decoded time frame to the rat's current movement direction every 5 ms. Theta sequences in both the linear track and open field tasks typically produced a virtual spatial path progressing from the rat's current location ahead of the animal during periods of active movement (Fig. 1A and figs. S1 and S2). However, theta oscillations rarely progressed uniformly in a single direction. They consisted of two distinct components, one that traveled ahead of the rat and a second that moved backward in the reverse direction of the animal's actual movement (Fig. 1, A and B, and figs. S1 and S2). The reverse component consistently occurred near the peak of theta oscillation (Fig. 1B), a time window associated with minimal hippocampal population activity (*6*). To ensure that the observation of a backward theta sequence was not a trivial result of low spike counts biasing the decoding algorithm, we initially restricted our analysis to the third of theta oscillations in each session with the highest firing rates within that window. On the basis of troughs in posterior probability distributions (Fig. 1B), we defined two windows in the theta oscillation: a forward window from phases 250° to 420°/60° and a reverse window from 80° to 230°, associated with statistically significant forward and backward virtual trajectories, respectively (Fig. 1C and figs. S3 and S4). Virtual representations of forward or reverse trajectories were observed within single theta oscillations (fig. S5) and were directly and independently correlated with population activity in the forward or reverse windows, respectively (fig. S6). The reverse component of the theta sequence was not a trivial consequence of our analysis methods or resetting of the forward trajectory (figs. S7 and S8).

Taking advantage of the fact that place cells display direction-selective firing patterns on linear tracks (*13*), we observed that both the forward and reverse components were predominantly encoded by neurons representing the rat's current movement direction (Fig. 1, D and E, and fig. S2). Thus, the reverse component of a theta sequence bears a direction-momentum relationship consistent with the rat running backward rather than turning around and running forward in the opposite direction. The direction-momentum relationship in reverse theta sequences thus bears a similarity to reverse replay observed within rest-based sharp-wave ripples (*1*, *2*).

Given that the reverse window was confined to phases of the theta oscillation associated with minimal firing across the hippocampal network, we investigated whether the backward portion of the theta sequence was encoded by a different population of neurons than the forward component—a population that might selectively fire while the remainder is relatively silent. Although most neurons (which we termed "unimodal cells") displayed a canonical (*6*) unimodal relationship between firing rate and theta phase, a subset of neurons ("bimodal cells") displayed a bimodal relationship with theta phase (Fig. 2, A and B, and fig. S9), consistent with the firing patterns reported for deep CA1 pyramidal neurons (*14*). The activity of bimodal cells across the theta oscillation was separated into two windows: a major peak between phases 200° and 430°/70°, comparable to the forward window, and a minor peak between 80° and 190°, consistent with the reverse window (Fig. 2B). Spatial representation during the reverse window, but not the forward window, was more strongly influenced by bimodal cell activity as opposed to unimodal activity (fig. S10). Bimodal cells were more likely to fire in ripples than unimodal cells (Fig. 2C) and further displayed an increased likelihood to participate in reverse replay as opposed to forward replay (Fig. 2D), indicating a neuronal link between reverse theta sequences and reverse replay. Bimodal cells had similar cluster quality measurements to those of unimodal cells (figs. S11 and S12) and were thus unlikely to be a trivial consequence of poor single-unit isolation. Similar to deep CA1 pyramidal neurons (*15*), bimodal cells were more likely to shift their preferred phase of firing within theta during REM (rapid eye movement) sleep (fig. S13). Furthermore, like deep CA1 neurons (*16*), bimodal cells displayed increased spatial information (fig. S12).

We next sought to identify the cellular mechanism that produces the reverse component of theta sequences. The predominant model accounting for the well-studied forward component of theta sequences is theta phase precession (*6*, *7*), the iterative firing of a place cell at progressively earlier phases of the theta oscillation as the rat passes through the neuron's firing field (*17*). We reasoned that the backward trajectory encoded within theta sequences may be explained by the opposite process: phase procession within individual place cells during the reverse window. When we examined the relationship of theta firing phase to the rat's location within each cell's place field, we identified clear phase precession during the major

[1]Neuroscience Graduate Program, Department of Neuroscience, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: brad.pfeiffer@utsouthwestern.edu

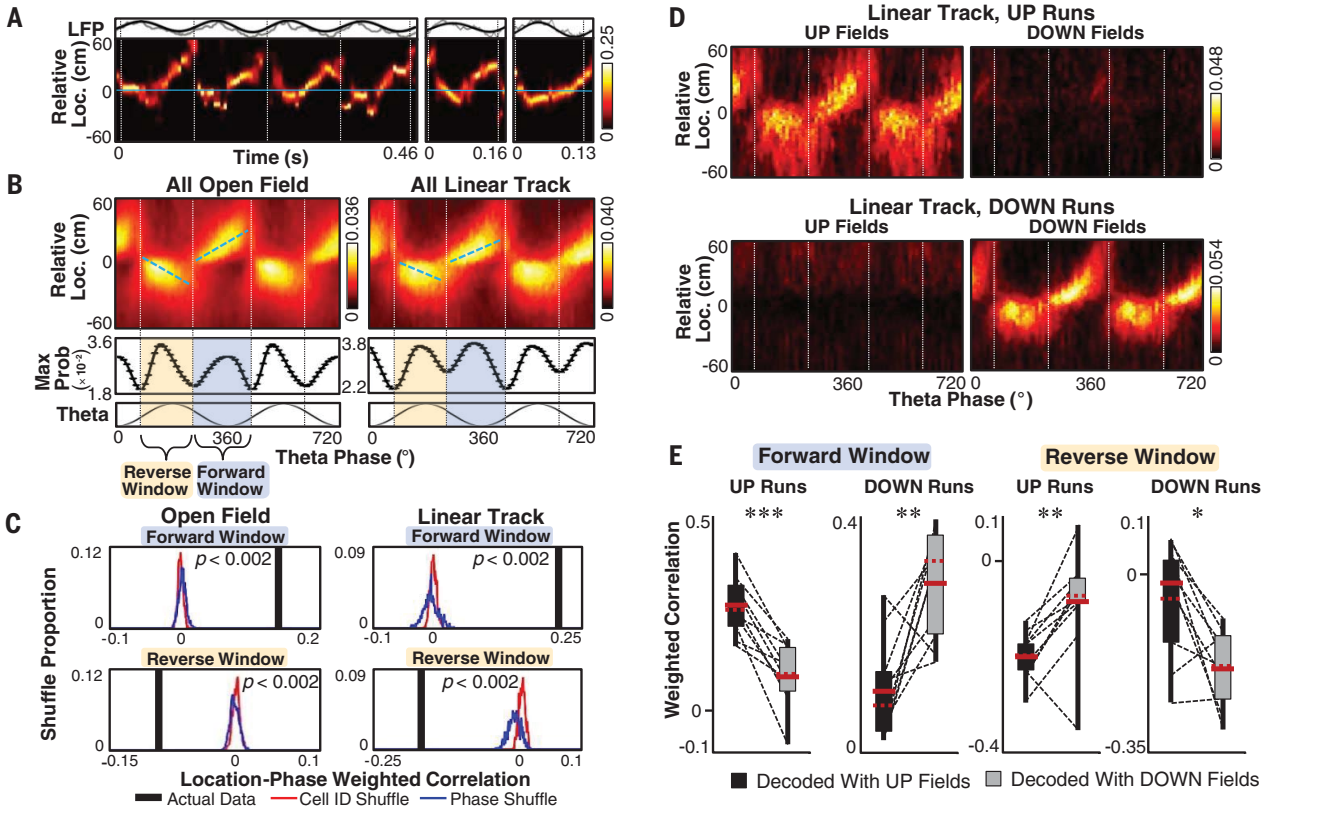


Fig. 1. Forward and reverse components of theta sequences. (**A**) Examples of theta sequences displaying both reverse and forward components. (Top) Raw (gray) and theta-filtered (black) local field potential (LFP) trace. Loc., location. (Bottom) Decoded spatial representation probability relative to the rat's current location (blue line) and movement direction (positive values). Dashed lines mark the 70° theta phase. (**B**) (Top) Across-session probability histogram of theta phase (10° bins) with maximum posterior probability at each position relative to rat position (2-cm bins) for the third of theta oscillations with highest firing near the peak of theta phase. Dashed blue lines indicate best fit (least-squares method). (Middle) Smoothed (Gaussian, σ = 10°) maximum value of each theta bin above. Vertical dashed lines indicate troughs in maximum probability (Max Prob) density (70° and 240°) used to define forward and reverse windows. (Bottom) Idealized theta oscillation. (**C**) Actual data and distribution of weighted correlation values for 500 theta phase or cell ID shuffles per theta oscillation in forward and reverse windows. The maximum Monte-Carlo *P* value for either shuffle displayed is noted. (**D**) As in the top portion of (B), but for only linear track sessions separated by runs across the track in the UP (top) or DOWN (bottom) direction, decoded with place fields calculated during up (left) or down (right) runs. (**E**) Per-session box-and-whisker plot (box, quartile; whiskers, extreme range), mean (red solid line), and median (red dashed line) of weighted correlations per session for forward and reverse windows during up or down runs decoded with up or down fields. Dashed lines represent individual session averages. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; Wilcoxon rank-sum test; *n* = 10 sessions.

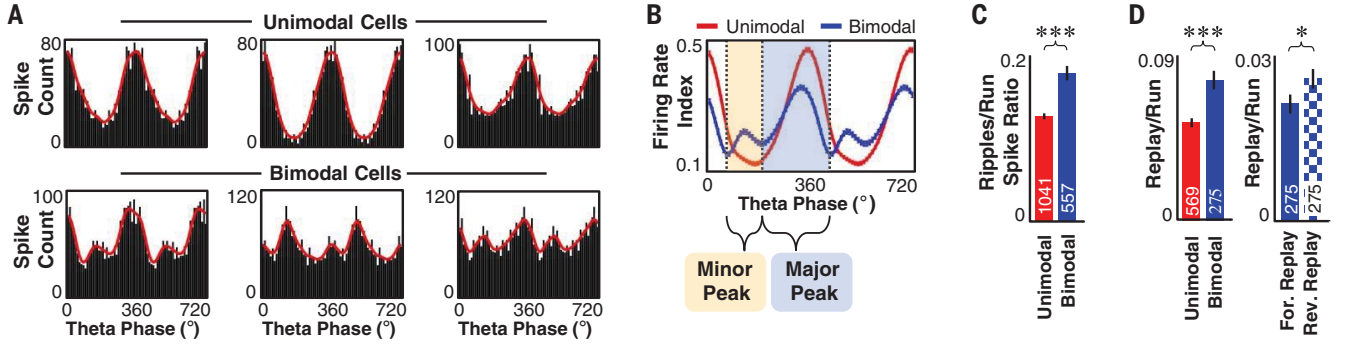


Fig. 2. Unimodal and bimodal cells. (**A**) Raw (black) and smoothed (red) histograms of action potential count per theta phase for example unimodal (top) and bimodal (bottom) cells. (**B**) Mean ± SEM firing rate index versus theta phase for all unimodal (red; *n* = 1041) and bimodal (blue; *n* = 557) cells across all open field and linear track sessions. Vertical lines mark troughs of mean firing for bimodal cells used to separate theta oscillations into major and minor peak windows. (**C**) Mean ± SEM ratio of spikes emitted across all ripples to spikes emitted during runs (velocity ≥10 cm/s) for unimodal or bimodal cells across all sessions. (**D**) (Left) Mean ± SEM ratio of spikes in significant replay events to spikes during runs for unimodal or bimodal cells across all linear track sessions. (Right) Mean ± SEM ratio of spikes in forward (solid) or reverse (checkered) replay to spikes during runs for bimodal cells across all linear track sessions. **P* < 0.05, ****P* < 0.001; Student's *t* test; cell number (*n*) is listed on the bar graphs.

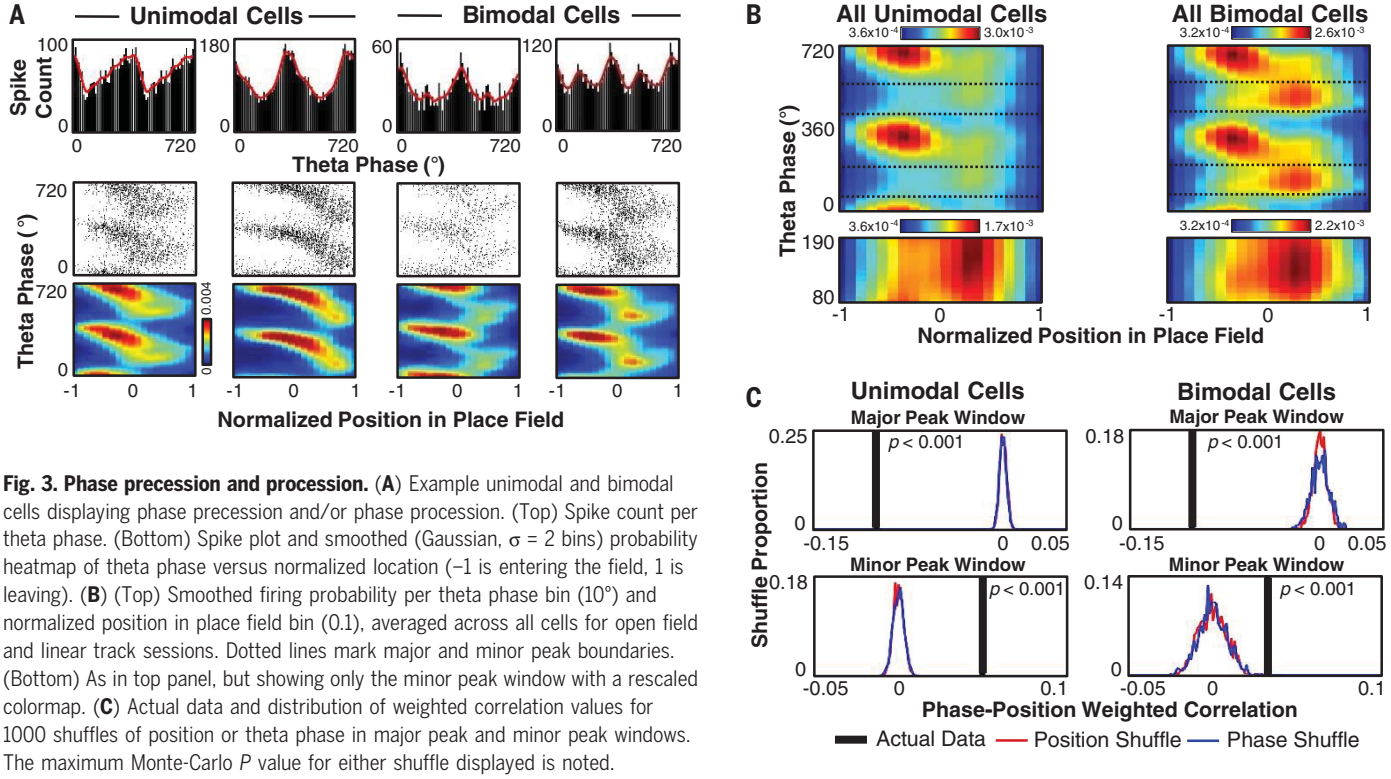


Fig. 3. Phase precession and procession. (**A**) Example unimodal and bimodal cells displaying phase precession and/or phase procession. (Top) Spike count per theta phase. (Bottom) Spike plot and smoothed (Gaussian, σ = 2 bins) probability heatmap of theta phase versus normalized location (−1 is entering the field, 1 is leaving). (**B**) (Top) Smoothed firing probability per theta phase bin (10°) and normalized position in place field bin (0.1), averaged across all cells for open field and linear track sessions. Dotted lines mark major and minor peak boundaries. (Bottom) As in top panel, but showing only the minor peak window with a rescaled colormap. (**C**) Actual data and distribution of weighted correlation values for 1000 shuffles of position or theta phase in major peak and minor peak windows. The maximum Monte-Carlo *P* value for either shuffle displayed is noted.

peak of activity for both unimodal and bimodal cells (Fig. 3, A and B, and fig. S14). However, during the minor peak of activity associated with the reverse window, many bimodal cells displayed significant phase procession (Fig. 3C and figs. S14 and S15). The average phase-versus-location plots for bimodal cells contained two distinct clusters rather than the single linear or curved relationship observed for unimodal neurons (Fig. 3B), as indicated in earlier work (*6*, *18*). Although their total firing rates were lower in the minor peak window, unimodal cells also displayed phase procession within this period (Fig. 3, B and C, and figs. S14 and S15). Thus, phase procession during the reverse window is not a characteristic aspect of bimodal cells but is likely a common feature of hippocampal activity during this time frame.

Finally, we sought to identify whether the forward and reverse theta sequences were driven by a common source or were instead regulated by independent inputs. Because forward and reverse sequences were respectively correlated with population activity during the forward and reverse windows (fig. S6), we reasoned that if a common input accounted for both the forward and reverse components, population activity in both windows would be strongly correlated. If forward and reverse theta sequences were instead driven by two distinct inputs, activity in these two windows should be independent. Neural activity within the major peak, but not the minor peak, positively correlated with the power of the hippocampal theta oscillation (Fig. 4 and fig. S16). Activity at the minor peak, but not the major peak, was correlated with the power of the beta oscillation (Fig. 4 and fig. S16). In addition, beta power was increased in theta oscillations expressing reverse theta sequences (fig. S17). Although some prior studies suggest that the hippocampus may receive both theta-frequency and beta-frequency inputs (*19*–*21*), several others indicate that two independent theta-frequency inputs that are roughly antiphase to one another drive activity in area CA1 (*14*, *22*–*24*), which can produce beta-frequency oscillatory activity. Our data are consistent with the second model (fig. S18) and support the hypothesis that activity in the forward and reverse windows is driven by two independent, antiphase theta-frequency inputs.

CA1 pyramidal neurons receive two primary, anatomically segregated excitatory inputs: the Shaffer collaterals originating in hippocampal area CA3 and the perforant path from layer 3 of the entorhinal cortex (EC3). The timing of these respective inputs is also segregated, with EC3 input arriving near the peak of CA1 theta oscillation (during the reverse window) and CA3 input arriving nearer the trough of CA1 theta oscillation (during the forward window) (*14*, *22*). Thus, the relative timing and independent expression that we observe in the forward and reverse components of theta sequences indicate that CA3 input selectively facilitates phase precession and drives the forward, prospective sequence, whereas antiphase EC3 input selectively facilitates phase procession and drives the reverse, retrospective sequence. Consistent with this interpretation, theta oscillations with significant reverse sequences have increased power in the fast and medium gamma bands (fig. S17), which is associated with EC input to CA1 (*14*, *25*), but not the slow gamma band, which is associated with CA3 input (*25*).

Although several models have been proposed to explain the observation of reverse replay (*1*, *26*, *27*), it remains unclear how reverse-ordered sequences can arise from forward-ordered activity. Given that activity within the reverse window is linked to participation in reverse replay events (Fig. 2D), our data suggest that synaptic inputs arriving during this window, likely originating from EC3, facilitate synaptic changes across the hippocampal network that underlie the expression of reverse replay. CA1 synapses receiving input from area CA3 have synaptic and plastic properties that are distinct from those of synapses receiving input from EC3 (*28*, *29*). The relative timing of EC3 input also affects plasticity at Shaffer collateral synapses (*30*, *31*). The increased likelihood of observing reverse replay immediately after a behavioral trajectory (*2*) and the lower level of reverse replay reported during sleep (*32*) suggest that plasticity generated in the reverse window may be stronger but more temporary than plasticity generated in the forward window. The power

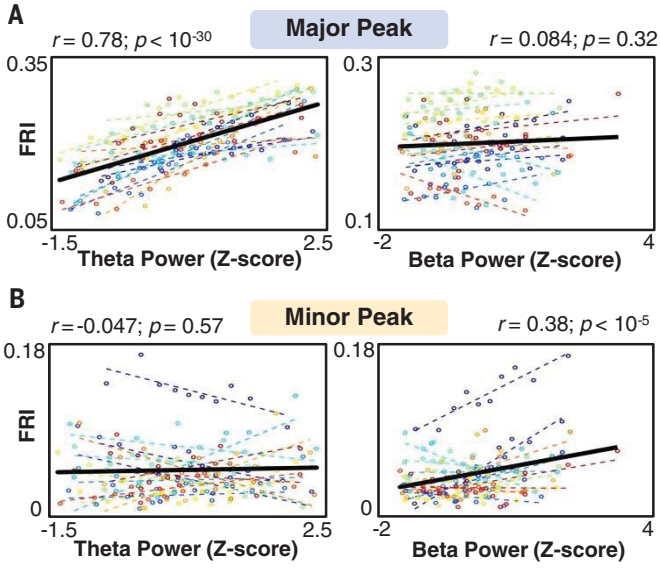


Fig. 4. Independent activity in forward and reverse components of theta sequences. For all open field and linear track sessions, firing rate index (FRI) in the major peak window (**A**) or minor peak window (**B**) as a function of theta (6 to 12 Hz) power (left) or beta (15 to 20 Hz) power (right), separated into 10 divisions per session. Colored lines indicate data and best fit for each session; black lines indicate best fit across all sessions. Statistics are based on repeated measures of correlation; n = 16 sessions.

of a beta-frequency oscillation, which we show to be correlated with the reverse theta sequence (Fig. 4 and figs. S16 and S17), is increased during early exploration of novel environments (*20*) and may represent enhanced EC3 input, which facilitates the rapid expression of reverse replay (*1*).

Although most studies of theta sequences have reported exclusively forward-ordered trajectories (*6*, *7*), a small number have suggested the presence of retrospective information encoded within some hippocampal theta oscillations (*8*, *32*–*34*). Most prior studies lacked the cell yield necessary to accurately quantify spatial sequences on a fine time scale. They were thus unable to observe the consecutive, phase-locked forward and reverse sequences we report here, particularly during the reverse window when population activity is low. Bimodal firing rates and theta phase–versus–location relationships of individual neurons have been previously reported (*6*, *14*, *18*). However, these earlier studies did not predict sequential representation of prior experience within individual theta oscillations. Finally, a recent study reported that the forward component of successive theta sequences alternates between possible future paths or heading directions during phases of theta that correspond to the forward window (*9*). Although these findings align with our observations, a notable contrast is that here we have specifically identified two distinct and independent information streams within each theta cycle: one that represents a possible future outcome and one that represents prior behavior in the reverse order. Our findings therefore facilitate understanding and integration of prior work into a broad, cohesive model of hippocampal circuit function, demonstrating alternating prospective and retrospective evaluation of behavior within individual theta oscillations.

REFERENCES AND NOTES

1. D. J. Foster, M. A. Wilson, *Nature* **440**, 680–683 (2006).
2. K. Diba, G. Buzsáki, *Nat. Neurosci.* **10**, 1241–1242 (2007).
3. G. E. Wimmer, Y. Liu, N. Vehar, T. E. J. Behrens, R. J. Dolan, *Nat. Neurosci.* **23**, 1025–1033 (2020).
4. D. Robbe, G. Buzsáki, *J. Neurosci.* **29**, 12597–12605 (2009).
5. K. A. Bolding, J. Ferbinteanu, S. E. Fox, R. U. Muller, *Hippocampus* **30**, 175–191 (2020).
6. W. E. Skaggs, B. L. McNaughton, M. A. Wilson, C. A. Barnes, *Hippocampus* **6**, 149–172 (1996).
7. G. Dragoi, G. Buzsáki, *Neuron* **50**, 145–157 (2006).
8. A. S. Gupta, M. A. van der Meer, D. S. Touretzky, A. D. Redish, *Nat. Neurosci.* **15**, 1032–1039 (2012).
9. K. Kay *et al.*, *Cell* **180**, 552–567.e25 (2020).
10. C. Drieu, R. Todorova, M. Zugaro, *Science* **362**, 675–679 (2018).
11. B. E. Pfeiffer, D. J. Foster, *Nature* **497**, 74–79 (2013).
12. B. E. Pfeiffer, D. J. Foster, *Science* **349**, 180–183 (2015).
13. B. L. McNaughton, C. A. Barnes, J. O'Keefe, *Exp. Brain Res.* **52**, 41–49 (1983).
14. A. Fernández-Ruiz *et al.*, *Neuron* **93**, 1213–1226.e5 (2017).
15. K. Mizuseki, K. Diba, E. Pastalkova, G. Buzsáki, *Nat. Neurosci.* **14**, 1174–1181 (2011).
16. N. B. Danielson *et al.*, *Neuron* **91**, 652–665 (2016).
17. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317–330 (1993).
18. Y. Yamaguchi, Y. Aota, B. L. McNaughton, P. Lipa, *J. Neurophysiol.* **87**, 2629–2642 (2002).
19. C. S. Lansink *et al.*, *J. Neurosci.* **36**, 10598–10610 (2016).
20. J. D. Berke, V. Hetrick, J. Breck, R. W. Greene, *Hippocampus* **18**, 519–529 (2008).
21. L. M. Rangel, A. A. Chiba, L. K. Quinn, *Front. Syst. Neurosci.* **9**, 96 (2015).
22. E. W. Schomburg *et al.*, *Neuron* **84**, 470–485 (2014).
23. M. Valero, L. M. de la Prida, *Curr. Opin. Neurobiol.* **52**, 107–114 (2018).
24. A. Navas-Olive *et al.*, *Nat. Commun.* **11**, 2217 (2020).
25. L. L. Colgin *et al.*, *Nature* **462**, 353–357 (2009).
26. R. A. Koene, M. E. Hasselmo, *Neural Netw.* **21**, 276–288 (2008).
27. T. Haga, T. Fukai, *eLife* **7**, e34171 (2018).
28. E. Arrigoni, R. W. Greene, *Br. J. Pharmacol.* **142**, 317–322 (2004).
29. A. Aksoy-Aksel, D. Manahan-Vaughan, *Front. Synaptic Neurosci.* **5**, 5 (2013).
30. H. Takahashi, J. C. Magee, *Neuron* **62**, 102–111 (2009).
31. J. Basu *et al.*, *Science* **351**, aaa5694 (2016).
32. A. M. Wikenheiser, A. D. Redish, *Hippocampus* **23**, 22–29 (2013).
33. K. W. Bieri, K. N. Bobbitt, L. L. Colgin, *Neuron* **82**, 670–681 (2014).
34. A. Cei, G. Girardeau, C. Drieu, K. E. Kanbi, M. Zugaro, *Nat. Neurosci.* **17**, 719–724 (2014).
35. M. Wang, D. Foster, B. Pfeiffer, Data from "Alternating sequences of future and past behavior encoded within hippocampal theta oscillations," Zenodo (2020); doi:10.5281/zenodo.3972156.

ACKNOWLEDGMENTS

Funding: This work was supported by NIH R01-MH085823, NIH R01-NS104829, the Alfred P. Sloan Foundation, and the Southwestern Medical Foundation. **Author contributions:** B.E.P. and D.J.F. conceived the original experiments. B.E.P. performed the original experiments and collected the original data. M.W. and B.E.P. conceived, designed, and performed the analyses. M.W. and B.E.P. wrote the manuscript. **Competing interests:** None declared. **Data and materials availability:** Data analysis code (MATLAB) is available on GitHub (github.com/Brad-E-Pfeiffer/ThetaForwardReverseCode) and Zenodo (*35*). All (other) data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6513/247/suppl/DC1
Materials and Methods
Figs. S1 to S18
Table S1
References (*36*–*38*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

21 February 2020; accepted 21 August 2020
10.1126/science.abb4151

Adeno-Associated Virus Titer Kit

Gyros Protein Technologies introduces the Gyrolab AAVX Titer Kit for adeno-associated virus (AAV) vector titer determination in AAV vector manufacturing. The kit for physical titer determination, one of the critical quality attributes of the U.S. Food and Drug Administration process validation guidelines, has a broad applicability across numerous AAV serotypes and adds to our wide range of ready-to-use kits that are already employed by scientists and bioengineers in the rapidly growing cell and gene therapy market, where AAV vectors are commonly used. The kit supports scientists in the development and production of cell and gene therapies to maximize performance and productivity, and reduces the time needed to produce patient treatments. With compressed development timelines for these therapies, the new kit generates results 4 times faster and requires 10 times less sample compared to traditional ELISA methods, accelerating AAV vector workflows and delivering high-quality data from small sample volumes. Featuring an analytical range of >3 logs, Gyrolab AAVX Titer Kit also minimizes repeat testing compared to ELISA.

Gyros Protein Technologies
For info: +46-(0)-18-56-63-00
www.gyrosproteintechnologies.com

RNA Extraction for COVID-19 Diagnostic Testing

Porvair Sciences announces a special version of its popular Homogenizer Spin Column optimized for RNA extraction from buccal swabs. A large number of current COVID-19 diagnostic tests are reliant on detection of viral RNA from buccal swabs by PCR. The column has been adapted to routinely improve yield and quality of viral RNA from nasal and throat swab samples. It provides a fast procedure for downstream RNA extraction, achieving great results in just minutes. The novel dual-frit extraction design reduces lysate viscosity and captures insoluble debris by centrifugation, so the homogenized lysate sample is ready for RNA extraction. Using the Homogenizer increases RNA yield as compared to alternative shredding columns or if no homogenization is undertaken. The column ensures faster homogenization than is possible with traditional methods using syringe and needles, with reduced waste and minimal loss of sample.

Porvair Sciences
For info: 800-552-3696
www.chromatrap.com/homogeniser-spin-column

Viral DNA/RNA Purification

Quick-DNA/RNA Viral MagBead kit is designed for high-throughput purification of viral DNA and/or RNA from plasma, serum, urine, cell culture media, blood, saliva, cellular suspensions, biopsies, and swab and fecal samples stored in our DNA/RNA Shield medium (for sample collection, nucleic acid preservation, and inactivation of pathogens). The kit also features a buffer system that facilitates complete viral-particle lysis for efficient nucleic acid isolation. Small (>50 nucleotide) and large (>200 kilobase) DNA and RNA are bound to magnetic beads, washed, and eluted. The isolated high-quality nucleic acids are ready for all downstream applications, such as next-generation sequencing, hybridization-based PCR detection, and quantitative reverse-transcription PCR detection.

Zymo Research
For info: 888-882-9682
www.zymoresearch.com/products/quick-dna-rna-viral-magbead

Automated Gene Synthesis

Codex DNA introduces the BioXp 3250 system, the next-generation, automated synthetic biology workstation for rapidly building DNA clones, variant libraries, gene fragments, and genomes.
Unlike traditional approaches that can take weeks or months to generate gene fragments, clones, or libraries, BioXp 3250 achieves these workflows in a single, overnight run. With increased gene-fragment lengths of up to 7,000 base pairs and an extensive selection of off-the-shelf vectors, BioXp 3250 offers scientists greater flexibility to generate full-length DNA constructs, clones, and libraries with unprecedented speed. In addition, the benchtop system allows laboratories to maintain complete control of intellectual property relating to their proprietary vectors. As a result, this single platform provides a secure method to exponentially accelerate discovery of candidate drugs for infectious disease, cancer, and answering gene-function questions.

Codex DNA
For info: 858-228-4115
codexdna.com/products/bioxp-system

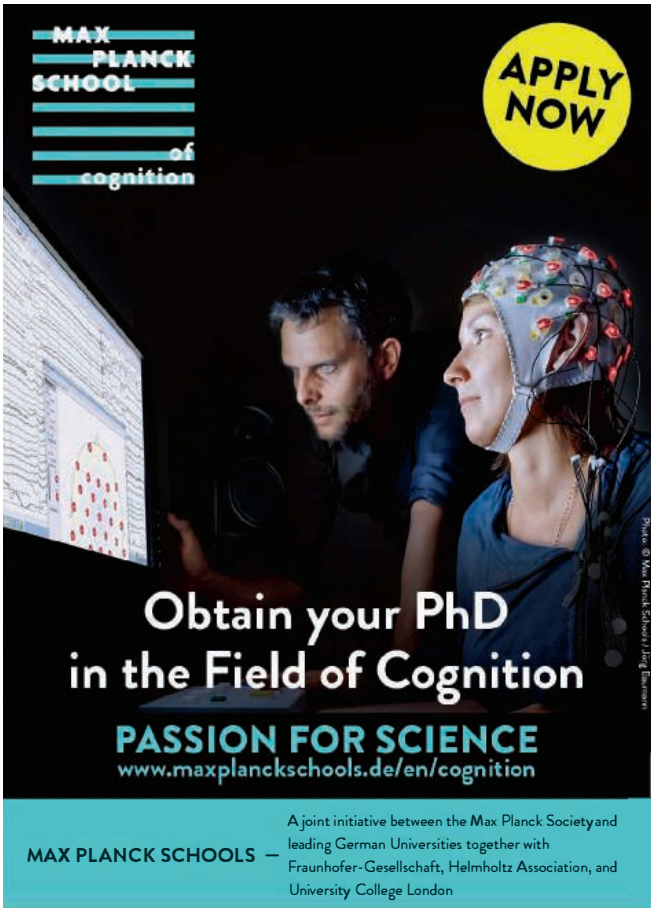
KAPA Target Enrichment

Roche Sequencing announces its KAPA Target Enrichment Portfolio, a new collection of powerful, hybridization-based target enrichment tools for clinical, translational, and basic research.
The portfolio includes (1) HyperDesign, a new, easy-to-use online tool for the design of custom probe panels that leverages Roche's renowned design expertise; (2) KAPA Target Enrichment Probes, high-performing probes that provide greater sequencing coverage, lower duplication rates, and improved coverage uniformity; (3) KAPA HyperChoice, fully customizable probe panels for human samples; (4) KAPA HyperExplore, fully customizable probe panels for nonhuman samples; (5) KAPA HyperExome, a compact whole-exome panel with broad coverage of human genomic and medical research databases (CCDS, RefSeq, Ensembl, GENCODE, ClinVar); and (6) HyperCap Workflow v3.0: a streamlined, automatable solution for integrated library preparation and target enrichment using KAPA HyperPrep or KAPA HyperPlus Kits, now updated for greater efficiency with KAPA Target Enrichment Probes.

Roche Sequencing
For info: 855-527-2246
www.roche.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.


MAX PLANCK SCHOOL of cognition
APPLY NOW
Obtain your PhD in the Field of Cognition
PASSION FOR SCIENCE
www.maxplanckschools.de/en/cognition
MAX PLANCK SCHOOLS –
A joint initiative between the Max Planck Society and leading German Universities together with Fraunhofer-Gesellschaft, Helmholtz Association, and University College London

COLUMBIA
VAGELOS COLLEGE OF PHYSICIANS AND SURGEONS

2021 Schaefer Research Scholars Program Awards

The Vagelos College of Physicians and Surgeons (VP&S) is pleased to announce the *2021 Schaefer Research Scholars Program Awards.* The *Awards,* made possible through a bequest from Dr. Ludwig Schaefer, are made annually to four research scientists who have distinguished themselves in human physiology, as broadly defined, and whose current work is of outstanding merit. The proposed research must have the potential to illuminate the field. Two awards are made to research scientists residing or working in North or South America, and two awards are made to research scientists residing or working outside North or South America. Each award consists of a $50,000 cash prize and up to $200,000 in direct research support.

Applications must include a cover sheet; research proposal (one page); a research budget (not to exceed $200,000 in total direct costs) delineated by cost category (salary, fringe, supplies, etc.) for one year (7/1/2021–6/30/2022); a curriculum vitae (not to exceed 10 pages); and a page summarizing applicant's research support. Internal candidates must obtain a nomination letter from their Department Chair. External candidates must present letters from the Dean or equivalent in their home institution as well as from the Columbia University Irving Medical Center collaborator, if applicable.

Nomination deadline: **November 19, 2020, at 5:00 p.m. (EST)**

To apply visit https://www.ps.columbia.edu/schaefer and submit all required materials.

Awardees will be notified in February 2021.

EPFL

Faculty Position in Interdisciplinary Cancer Research

at the Ecole polytechnique fédérale de Lausanne (EPFL)

The School of Life Sciences of EPFL (Ecole polytechnique fédérale de Lausanne) invites applications for a tenure track assistant professor position in the field of **Interdisciplinary Cancer Research**.

This search is part of major initiatives in the Lake Geneva region to promote cancer research, which is increasingly driven by integrating research from different fields ranging from basic sciences (life sciences, physics, chemistry and engineering) to clinical research and treatment.

The successful candidate will: join the faculty of the Swiss Institute for Experimental Cancer Research (ISREC) http://isrec.epfl.ch; develop an independent internationally prominent research program in the broad domain of interdisciplinary cancer research and its potential therapeutic applications; participate in both undergraduate and graduate teaching; and supervise PhD students and postdoctoral fellows. Candidates may work in any cancer-related area, including cancer genetics, functional genomics and genome instability, epigenetic regulation of cancer genotypes, cancer metabolism, computational cancer genetics, bioengineering of cell-based therapies and oncolytic viruses, and synthetic biology, systems biology, chemical biology of cancer. Expertise in data science is also encouraged.

The successful candidate will be a part of EPFL's cancer research institute (ISREC), and is expected to perform and coordinate highly interactive biomedical research, reaching out and taking advantage of EPFL's interdisciplinary campus (Schools of Basic Sciences, Engineering, and Information and Communication Technologies) and its involvement in the multi-institutional Swiss Cancer Center Leman, which brings together EPFL, the Universities of Lausanne and Geneva, and clinical and translational research components of the University Hospitals of Lausanne and Geneva.

Applications should include a cover letter, a curriculum vitae, a list of publications (annotated to indicate the candidate's contributions) a synopsis of major accomplishments, and a concise statement of future research agenda and teaching interests, along with the contacts of 3-5 referees who can provide letters of recommendations. Applications should be uploaded as PDF files to the recruitment web site:

https://facultyrecruiting.epfl.ch/position/23691279

Formal evaluation of candidates will begin on **January 8, 2021**, and continue until the position is filled.

Enquiries may be sent to:
Prof. Freddy Radtke
Search Committee Chair
E-mail: cancer.research@epfl.ch

For additional information on EPFL, the ISREC institute and the school of life sciences, please consult: www.epfl.ch, sv.epfl.ch

EPFL is an equal opportunity employer and family friendly university. It is committed to increasing the diversity of its faculty. It strongly encourages women to apply.

Penn Engineering
University of Pennsylvania

ESE Tenured or Tenure-track Faculty Openings, 2020 - 2021

The School of Engineering and Applied Science at the University of Pennsylvania is growing its faculty by 33% over a five-year period. As part of this initiative, the **Department of Electrical and Systems Engineering** is engaged in an aggressive, multi-year hiring effort for multiple tenure-track positions at all levels. Candidates must hold a Ph.D. in Electrical Engineering, Computer Engineering, Systems Engineering, or related area. The department seeks individuals with exceptional promise for, or proven record of, research achievement, who will take a position of international leadership in defining their field of study and who will excel in undergraduate and graduate education. Leadership in cross-disciplinary and multi-disciplinary collaborations is of particular interest. We are interested in candidates in all areas that enhance our research strengths in:

- **Nanodevices and nanosystems** (nanoelectronics, MEMS/NEMS, power electronics, nanophotonics, nanomagnetics, quantum devices, integrated devicesand systems at nanoscale); **https://apptrkr.com/2007344**
- **Circuits and computer engineering** (analog, RF, mm-wave, digital circuits, emerging circuit design, computer engineering, IoT, beyond 5G, andcyber-physical systems); **https://apptrkr.com/2007395**
- **Information and decision systems** (control, optimization, robotics, data science, machine learning, communications, networking, information theory, signal processing). **https://apptrkr.com/2008072**

Prospective candidates in all areas are strongly encouraged to address large-scale societal problems in energy, transportation, health, agriculture, food and water, economic and financial networks, social networks, critical infrastructure, and national security. We are especially interested in candidates whose interests are aligned with the school's strategic plan, https://www.seas.upenn.edu/about/penn-engineering-2020/

Diversity candidates are strongly encouraged to apply. Interested persons should submit an online application by following the links above and include curriculum vitae, research, teaching, and diversity statements, and at least three references. Review of applications will begin on January 4, 2021.

The University of Pennsylvania values diversity and seeks talented students, faculty and staff from diverse backgrounds. The University of Pennsylvania is an equal opportunity and affirmative action employer. Candidates are considered for employment without regard to race, color, sex, sexual orientation, gender identity, religion, creed, national or ethnic origin, citizenship status, age, disability, veteran status or any other legally protected class.

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers AAAS In partnership with:

UCSF University of California San Francisco
UIC
Burroughs Wellcome Fund
UMASS University of Massachusetts Medical School
FASEB Federation of American Societies for Experimental Biology

DEPARTMENT OF BIOTECHNOLOGY
MINISTRY OF SCIENCE & TECHNOLOGY
GOVERNMENT OF INDIA

सत्यमेव जयते

United Nations Educational, Scientific and Cultural Organization

रीजनल सेंटर फॉर बायोटेक्नोलॉजी
Regional Centre for Biotechnology

RAMALINGASWAMI RE-ENTRY FELLOWSHIP: 2020-2021

Regional Centre for Biotechnology (RCB), on behalf of the Department of Biotechnology (DBT), Ministry of Science & Technology, Government of India, invites applications from Indian Nationals working in overseas research institutions for the "Ramalingaswami Re-entry Fellowship" to provide them an attractive avenue for pursuing their R&D interests in Indian institutions. Those who have returned to India within one year of the closing date of this advertisement are also eligible to apply.

APPLICATION CLOSING DATE: 31.10.2020

Further details on the fellowship, eligibility criteria, and the application format can be downloaded from RCB or DBT website (www.rcb.res.in, www.dbtindia.gov.in). Applications in the prescribed format, duly forwarded by the host institution, may be sent as a hard copy to The Executive Director, Regional Center of Biotechnology, NCR Biotech Science Cluster, Faridabad-Gurugram Expressway, Faridabad - 121001, India, and mailed as a single PDF to dbt.rlsfellowship@rcb.res.in to reach by 31st October, 2020.

UTSouthwestern Medical Center

FACULTY POSITION
BACTERIAL PATHOGENESIS.

The Department of Microbiology at UT Southwestern Medical Center is seeking a new faculty member in bacterial pathogenesis at the **Assistant Professor** (tenure track) level. Appointment rank will be commensurate with academic accomplishments and experience. The appointee will be expected to develop a front-rank, competitive, independent research program on a medically relevant bacterial pathogen(s) and/or on concepts relevant to the human microbiome. An important academic responsibility also will be the instruction and mentoring of graduate students. An attractive start-up package, including a competitive salary and generous laboratory space in a modern building, is available to conduct research within a highly dynamic environment of a leading medical microbiology department (**https://www.utsouthwestern.edu/education/medical-school/departments/microbiology**).

Candidates will be considered for our $2M Endowed Scholars (start-up) Program (**http://www.utsouthwestern.edu/education/programs/nondegree-programs/other-programs/endowed-scholars/index.html**).

Candidates should have a Ph.D. and/or M.D. degree with at least 3-4 years of postdoctoral experience and an exceptional publication record. Please send a cover letter, C.V., contact information for three letters of recommendation, and a brief summary of future research to: **BacterialPathogenesisSearchCommittee@utsouthwestern.edu**.

UT Southwestern Medical Center is an Affirmative Action/Equal Opportunity Employer. Women, minorities, veterans and individuals with disabilities are encouraged to apply.

UTSouthwestern Medical Center

TENURE-TRACK POSITION

The Department of Physiology invites outstanding scientists with Ph.D., M.D., or equivalent degrees to apply for tenure-track faculty positions at the level of **Assistant Professor**.

Candidates who bring innovative approaches to the study of any under-explored/unexplored questions broadly related to physiology are encouraged to apply. The scientific excellence of the candidates is more important than the specific area of research. These positions are part of the continuing growth of the Department at one of the country's leading academic medical centers. They will be supported by significant laboratory space, competitive salaries, state-of-the-art core facilities and exceptional start-up packages. The University of Texas Southwestern Medical Center is the scientific home to six Nobel Prize laureates and many members of the National Academy of Sciences and Institute of Medicine. UT Southwestern conducts more than 3,500 research projects annually totaling more than $417 million. Additional information about the Department of Physiology can be found at **http://www.utsouthwestern.edu/education/medical-school/departments/physiology/index.html**.

Information regarding careers can be found at:
https://jobs.utsouthwestern.edu/.

Applicants should submit a CV, a brief statement of current and proposed research, and a summary of your two most significant publications describing the importance of the work (100-150 words each). Please arrange to have three letters of recommendation sent on his/her behalf. All items should be submitted to: **http://academicjobsonline.org/ajo/jobs/16617**. Completed applications will be reviewed starting **November 1, 2020**. You may email questions to **ron.doris@utsouthwestern.edu**.

UT Southwestern Medical Center is an Equal Opportunity/Affirmative Action Employer. Women, minorities, veterans and individuals with disabilities are encouraged to apply.

Science Careers online @sciencecareers.org

By **Sergio E. Ramos**

A pandemic journey

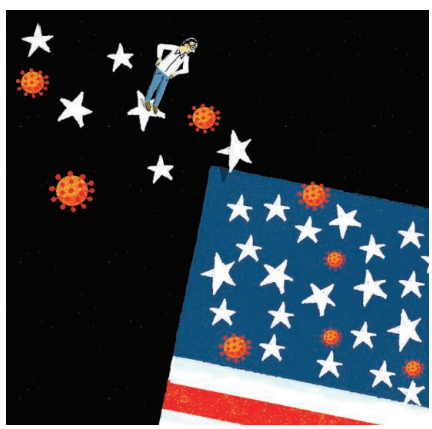
"There's no need to do a test; you have all the symptoms," the doctor said. "You have COVID-19." It was June, and I was in my home country of Mexico after spending 3 months stuck in Switzerland. I had hoped to return to the United States, where I am a postdoc, but as a noncitizen, I wasn't allowed to re-enter from Europe. I was told my best shot at getting back into the country might be through Mexico—where the pandemic was just ramping up. Between the virus and the U.S. government's anti-immigrant tone, I felt anxious and frustrated. I was starting to question whether the advantages of being an international scientist outweighed the drawbacks.

Six years earlier, I had left Mexico to pursue my Ph.D. in Europe. I knew the move would bring countless career benefits, including improving my English and having the opportunity to perform world-class science in an international environment. As I finished my degree, my Ph.D. adviser encouraged me to pursue training in the United States to further broaden my scientific horizons. That would mean moving away from my partner—a Ph.D. student in Switzerland—but we figured we could visit each other regularly and would manage to be in the same place in a few years.

In March, I was excited to spend a week back in Switzerland, meeting with collaborators and seeing my partner. COVID-19 was on my radar, and I used a mask and hand sanitizer regularly. Then, halfway through the week, the U.S. government announced that noncitizens coming from much of Europe would not be allowed to enter the country—starting 3 days before I was scheduled to return.

The ban was only supposed to last for 30 days, and my advisers were fine with me working remotely, so I changed my flight accordingly. I kept busy with reading, writing, reviewing, and organizing my summer field experiments, assuming that I would be back in the United States and everything would be normal again.

But as time wore on, it became clear that things would not be going back to normal any time soon. My field experiments were canceled. I called various government offices to ask when the ban would be lifted, but nobody knew. Still, I needed to get back eventually—after all, I had only brought 1 week's worth of belongings with me. An official suggested that a route via Canada or Mexico was my only option, though I'd need to quarantine for 14 days in either country before I would be allowed to re-enter the United States.



"Moving around the world has offered a lot of advantages, but it also comes with costs."

Mexico was the obvious choice, but I thought it was an absurd idea. I would have to quarantine with my family, and getting to them would require taking three flights and passing through four airports. Why risk exposing myself to the virus and endangering my relatives?

But in June, I got an email from my university's office of international services that forced my decision. It warned of the likely release of a White House executive order that would block some work and student visa holders from reentering the country. The email encouraged any recipients who were abroad to return as soon as possible. Less than a week later, I was flying to Mexico.

Five days after my arrival, my parents, brothers, and I started to feel sick with fevers, weakness, and joint pain. By the second week, the original symptoms disappeared, but I couldn't smell or taste anything, and my lungs felt odd. Luckily, our cases were relatively mild. After about a month, I was recovered enough to safely travel again.

When I finally got back to the United States, I couldn't help but ask myself, "Was all this worth it?" I'm now about 1 year into what is supposed to be a 2-year position. But depending on what happens with the pandemic and the U.S. election, I may decide to leave early. I'm not sure exactly where I would go, but I don't want to stay in a country where I am not valued despite working, paying my taxes, and fulfilling my responsibilities to my community. Moving around the world has offered a lot of advantages, but it also comes with costs. These days, those costs may have become too high. ■

Sergio E. Ramos is a postdoc at the University of Pittsburgh in Pennsylvania. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS